

# Automorphic Number Theory

Henryk Iwaniec

## CONTENTS

1. About these lectures	35
2. Classical Automorphic Forms	37
3. Maass Forms - Spectral Theory	41
4. Orthogonality of Hecke Eigenvalues	47
5. Orthogonality of Elliptic Curves	50
References	52

### 1. About these lectures

Number theory and automorphic forms cannot be separated nowadays in modern mathematics, although from a historical perspective both subjects grew in their own way for some time. Since number theory came first, I tend to think of automorphic forms as tools to serve the arithmetic. This attitude will be apparent during my lectures. Of course, with the advances in the first field, the second one benefits as well, that is to say number theory also offers new tools for general theory of automorphic forms.

I will not talk about geometric aspects of automorphic forms nor their connections with physics, representation theory. However, I shall emphasize their role in harmonic analysis because they are indispensable in analytic number theory, which is my primary subject of interest. A lot has been presented to the general forum about modular forms in algebraic number theory, in particular after resolution of the Fermat last theorem [W], [BCDT]. Therefore I shall limit my venture to the topics which help to grasp the essence of modularity in number theory as a whole.

The results shown during these lectures are selected for illustration of great ideas and to articulate the state of the particular subjects, so they are not always the sharpest or the deepest up to date accomplishments. These notes were prepared just for the audience of my talk, so they must be regarded as a draft for a more substantial exposition, which unfortunately was never written. The author expresses his apology for not fulfilling his promise.

## 2. Classical Automorphic Forms

These are generalizations of periodic functions. Let

$$\Gamma = \Gamma_0(q) = \left\{ \begin{pmatrix} ab & \\ & cd \end{pmatrix} \in SL_2(\mathbb{Z}); \quad c \equiv 0 \pmod{q} \right\}$$

and  $\chi \pmod{q}$ , a Dirichlet character. Let  $k \geq 0$  be an integer with

$$\chi(-1) = (-1)^k.$$

A function  $f$  on the upper half-plane

$$\mathbb{H} = \{z = x + iy; \quad x \in \mathbb{R}, \quad y \in \mathbb{R}^+\}$$

is said to be modular of level  $q$ , character  $\chi$ , and weight  $k$  if  $f(z)$  is holomorphic and it satisfies the automorphy equations

$$f\left(\frac{az + b}{cz + d}\right) = \chi(d) (cz + d)^k f(z)$$

for any  $\gamma = \begin{pmatrix} ab \\ cd \end{pmatrix} \in \Gamma$ . Moreover  $f(z)$  is required to be holomorphic at every cusp. In particular, for the cusp at infinity this condition means that  $f(z)$  has the Fourier expansion

$$f(z) = \lambda_f(0) + \sum_1^{\infty} \lambda_f(n) n^{\frac{k-1}{2}} e(nz), \quad e(z) = e^{2\pi iz}.$$

Note we pull out the factors  $n^{\frac{k-1}{2}}$  to make  $\lambda_f(n)$  nicer for analysis. The Eisenstein series and the Poincaré series are basic examples:

$$P_m(z) = \sum_{\gamma \in \Gamma_\infty \setminus \Gamma} \chi(d)(cz + d)^{-k} e(m\gamma z), \quad m = 0, 1, 2, \dots,$$

but not really significant for arithmetic. More important in this regard are, for example, the theta series

$$\theta_{PQ}(z) = \sum_{m \in \mathbb{Z}^r} P(m) e(Q(m)z).$$

They are associated with a positive definite quadratic form  $Q(x) = \frac{1}{2} {}^t x A x$  and harmonic polynomials  $P(x)$ . The modularity of  $\theta_{PQ}(z)$  and estimates for its Fourier coefficients (Ramanujan's conjecture holds due to P. Deligne only if  $r$  is even), leads to the equidistribution of integral points on ellipsoids (see results for any  $r \geq 3$  in Duke [D1] and Duke-Schultze-Pillot [DS-P]). If  $f$  has no zero terms at every cusp, it is called a cusp form. The linear space  $S_k(\Gamma, \chi)$  of cusp forms is a finite dimensional Hilbert space with the inner product

$$\langle f, g \rangle = \int_{\Gamma \setminus \mathbb{H}} f(z) \bar{g}(z) y^k d\mu z$$

where  $\Gamma \backslash \mathbb{H}$  is a fundamental domain and  $d\mu z = y^{-2} dx dy$  is the  $SL_2(\mathbb{R})$  invariant measure. This structure makes the space  $S_k(\Gamma, \chi)$  very wealthy, not only for analytic considerations, but it also plays a role in higher arithmetic. First of all, it allows one to select primitive forms (also called newforms, which are eigenfunctions of all the Hecke operators

$$(T_n f)(z) = \frac{1}{\sqrt{n}} \sum_{ad=n} \chi(a) \left(\frac{a}{d}\right)^{k/2} \sum_{0 \leq b < d} f\left(\frac{az+b}{d}\right).$$

Let

$$T_n f = \lambda_f(n) f, \quad \text{for all } n \geq 1.$$

With proper normalization, the eigenvalues  $\lambda_f(n)$  agree with the Fourier coefficients, i.e.

$$f(z) = \sum_1^{\infty} \lambda_f(n) n^{\frac{k-1}{2}} e(nz).$$

The significance of primitive forms is most visible in the multiplicativity of the coefficients

$$\lambda_f(m) \lambda_f(n) = \sum_{d|(m,n)} \chi(d) \lambda_f\left(\frac{mn}{d^2}\right).$$

Better yet, this property transcends in the Euler product for the associated  $L$ -function

$$L(f, s) = \sum_1^{\infty} \lambda_f(n) n^{-s} = \prod_p (1 - \lambda_f(p) p^{-s} + \chi(p) p^{-2s})^{-1}.$$

Why does a God given  $L$ -function (of degree two) correspond to a primitive cusp form? Well, because it is defined by an Euler product at first place. Starting from an Euler product the problem of analytic continuation is paramount. It shows somehow that the eigenvalues  $\lambda_f(p)$  at different places see each other and are equidistributed to cause a considerable cancellation. For example, the analytic continuation of the Artin  $L$ -functions

$$L(\rho, s) = \prod_{\mathfrak{p} \text{ unram}} \det \left( 1 - \rho \left( \left[ \frac{K/\mathbb{Q}}{\mathfrak{p}} \right] \right) (N\mathfrak{p})^{-s} \right)^{-1} L_{\text{ram}}(\rho, s)$$

yields (essentially) the reciprocity law.

Harder cases are the  $L$ -functions defined geometrically, the most successful example being the Hasse-Weil  $L$ -function of an elliptic curve

$$E : y^2 = x^3 + ax + b, \quad \Delta = -16(4a^3 + 27b^2) \neq 0,$$

the modularity of which is among the greatest achievements of mathematics in the past century (see [W], [BCDT]).

Simpler, but still attractive, examples of primitive cusp forms are those associated with quadratic number fields  $K = \mathbb{Q}(\sqrt{D})$ . Suppose  $K$  is imaginary, i.e. its discriminant  $D < 0$ . Let  $\psi$  be a character of the class group  $H = I/P$  and

$$\lambda_\psi(n) = \sum_{N\mathfrak{a}=n} \psi(\mathfrak{a}).$$

Then the  $L$ -function

$$\begin{aligned} L(s, \psi) &= \sum_1^\infty \lambda_\psi(n) n^{-s} \\ &= \sum_{\mathfrak{a}} \psi(\mathfrak{a})(N\mathfrak{a})^{-s} = \prod_{\mathfrak{p}} (1 - \psi(\mathfrak{p})(N\mathfrak{p})^{-s})^{-1} \end{aligned}$$

corresponds to a modular form  $f_\psi(z)$  of weight  $k = 1$ , level  $q = |D|$  and character  $\chi = \chi_D$  (the Kronecker symbol). If  $\psi$  is not real, then

$$f_\psi(z) = \sum_1^\infty \lambda_\psi(n) e(nz) \in S_1(\Gamma_0(|D|), \chi_D).$$

In general, how large is the space  $S_k(\Gamma_0(q), \chi)$  in terms of the level  $q$ ? If  $k \geq 2$ , we know that

$$\dim S_k(\Gamma_0(q), \chi) \asymp kq;$$

but for  $k = 1$ , we see an entirely different picture.

All the weight-one primitive cusp forms are presumably associated with two-dimensional Galois representations and conversely (see the works of Deligne-Serre [DS] and Serre [Se]). If  $q$  is prime (necessarily  $q \equiv 3 \pmod{4}$ ), note that the complex Hecke characters  $\psi$  of the ideal class group of  $K = \mathbb{Q}(\sqrt{-q})$  yield exactly  $\frac{1}{2}(h(-q) - 1)$  distinct cusp forms, whereas the conjecture of J-P. Serre [Se] asserts there are at most  $O(q^\epsilon)$  other cusp forms, i.e.

$$\dim S_1(\Gamma_0(q), \chi_{-q}) = \frac{1}{2}(h(-q) - 1) + O(q^\epsilon) \ll q^{\frac{1}{2}} \log q.$$

Using a trace formula, it is not difficult to estimate the dimension of  $S_1(\Gamma_0(q), \chi_{-q})$  by  $O(q/\log q)$ .

Why is any better bound hard to get? A comprehensive answer can be given only by considering the whole spectrum of cusp forms, including the real-analytic ones of Maass. The reason is that  $S_1(\Gamma_0(q), \chi_{-q})$  is just one eigenspace of the Laplace operator  $\Delta_1$  with eigenvalue  $\lambda_1 = \frac{1}{4}$ , which is not isolated (the bottom of the continuous spectrum). Consequently, any attempt to catch this part exclusively by playing with

test functions in relevant trace formula is doomed to failure due to the uncertainty principle of harmonic analysis. The arithmeticity of the weight one cusp forms must be exploited. In view of these intrinsic barriers, the following estimate of W. Duke [D2] came as a surprise:

$$\dim S_1(\Gamma_0(q), \chi_{-q}) \ll q^{\frac{1}{12}}.$$

Duke's work is a spectacular example of how arithmetical arguments enhance the analytic ones. He makes best use of the new technology — the amplification method. A thorough presentation of this method is given by P. Michel [M2].

### 3. Maass Forms - Spectral Theory

The completeness of a space in which to operate is an indispensable working condition in analysis. The Hilbert space of automorphic forms is where the modern analytic number theory works most effectively.

Consider functions  $F : \mathbb{H} \rightarrow \mathbb{C}$  which transform by

$$F\left(\frac{az+b}{cz+d}\right) = \chi(d) \left(\frac{cz+d}{|cz+d|}\right)^k F(z)$$

for any  $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$ . This is like in classical modular forms (periodicity, or symmetry), however the holomorphy condition is replaced by the requirement that  $F$  satisfies the second order partial differential equation

$$(\Delta_k + \lambda)F = 0.$$

Here  $\lambda = \lambda(s) = s(1-s) \in \mathbb{C}$  is a complex number - an eigenvalue of the Laplace - Beltrami operator

$$\Delta_k = y^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right) - ik y \frac{\partial}{\partial x}.$$

Because  $\Delta_k$  is elliptic, its eigenfunctions  $F = F(z) = F(x, y)$  are real-analytic in  $x, y$ . These are called Maass forms after [M].

We are interested in the  $L^2$ -space of automorphic functions with the inner product

$$\langle F, G \rangle = \int_{\Gamma \backslash \mathbb{H}} F(z) \overline{G(z)} d\mu z$$

in which  $\Delta_k$  has self-adjoint extension and is bounded below, precisely

$$\langle F, -\Delta_k F \rangle \geq \frac{|k|}{2} \left( 1 - \frac{|k|}{2} \right) \langle F, F \rangle.$$

The Fourier series expansion of a Maass form is a little bit more complicated than that of the classical forms, the exponential function being replaced by a Whittaker function. Nevertheless the theory of Hecke operators is pretty much the same. A primitive Maass cusp form  $u_j(z)$  of the Laplace eigenvalue  $\lambda_j = s_j(1-s_j) = \frac{1}{4} + t_j^2$  is a common eigenfunction of all the Hecke operators  $T_n$ , say

$$T_n u_j = \lambda_j(n) u_j, \quad n = 1, 2, 3, \dots$$

(the operators  $T_n$  are given by the same formula as for the holomorphic forms, but with  $k$  replaced by zero due to our normalization).

A. Selberg [S] established the complete spectral resolution of  $\Delta_k$ . He showed that besides the infinite dimensional discrete spectrum spanned

by Maass cusp forms, there is a continuous spectrum spanned by the Eisenstein series on the critical line. The presence of the continuous spectrum obscures the picture, however there are examples when the Eisenstein series do help to derive sharp estimates for the classical  $L$ -functions. Let me quote from [CI] the following

$$L(s, \chi_q) \ll |s|^{2003} q^{\frac{1}{6} + \varepsilon}, \text{ if } \operatorname{Re} s = \frac{1}{2},$$

the implied constant depending only on  $\varepsilon$ . Perhaps we should tell that this estimate (which improves the celebrated bound  $|s|q^{\frac{3}{16} + \varepsilon}$  of D. Burgess in the  $q$ -aspect) uses the Riemann Hypothesis for varieties over a finite field (proved by P. Deligne).

To see the position of the classical holomorphic cusp forms in the whole spectrum of Maass forms, it is necessary to apply two differential operators of order one

$$K_k = \frac{k}{2} + y \left( i \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \right)$$

$$\Lambda_k = \frac{k}{2} + y \left( i \frac{\partial}{\partial x} - \frac{\partial}{\partial y} \right).$$

The first operator  $K_k$  raises the weight of forms by 2, and the second operator  $\Lambda_k$  reduces the weight of forms by 2. Note that

$$\Delta_k = -K_{k-2}\Lambda_k - \lambda\left(\frac{k}{2}\right) = -\Lambda_{k+2}K_k - \lambda\left(\frac{-k}{2}\right).$$

Moreover if  $F, G, \Delta_k F, \Delta_k G$  are bounded, then

$$\begin{aligned} \langle F, -\Delta_k G \rangle &= \langle K_k F, K_k G \rangle + \lambda\left(\frac{-k}{2}\right) \langle F, G \rangle \\ &= \langle \Lambda_k F, \Lambda_k G \rangle + \lambda\left(\frac{k}{2}\right) \langle F, G \rangle. \end{aligned}$$

In particular, for  $F = G$ , we see that  $\lambda\left(\frac{|k|}{2}\right) = \frac{|k|}{2}\left(1 - \frac{|k|}{2}\right)$  is indeed the bottom of the spectrum as said earlier. For  $k = 1$ , this is at  $\lambda_1 = \frac{1}{4}$ . Moreover it follows that  $K_k, \Lambda_k$  operate on eigenspaces of  $\Delta_k$  isometrically except for the lowest one, which is annihilated. These, the lowest Maass cusp forms, say  $F(z)$ , correspond to the classical holomorphic cusp forms  $f(z)$  by the formula

$$f(z) = y^{\frac{k}{2}} F(z), \quad (\Delta_k + \lambda\left(\frac{k}{2}\right)) F = 0.$$

It was not known to H. Maass whether his space of real-analytic cusp forms, say  $C_k(\Gamma, X)$ , has infinite dimension. To answer this question, A. Selberg [S] developed the trace formula and his elegant zeta function by means of which he established the following asymptotic formula for the discrete spectrum (the so called Weyl's law). Put

$$(\Delta_k + \lambda_j)u_j = 0, \quad \lambda_j = \lambda_j(s_j) = s_j(1 - s_j), \quad s_j = \frac{1}{2} + it_j.$$



Then we have

$$\#\{j; 0 < t_j \leq T\} \sim \frac{1}{4\pi} \text{vol}(\Gamma \backslash \mathbb{H}) T^2, \text{ as } T \rightarrow \infty,$$

where

$$\text{vol}(\Gamma \backslash \mathbb{H}) = \frac{\pi}{3} q \prod_{p|q} \left(1 + \frac{1}{p}\right).$$

The Phillips-Sarnak work [PS] on spectral deformations gives strong evidence that the abundance of cusp forms is the feature (a treasure) of arithmetic groups, not a general rule. Hence it is very surprising and frustrating that nobody has ever constructed a Maass cusp form of weight-zero for the full modular group  $SL_2(\mathbb{Z})$ , in spite of the fact that there are infinitely many of them by counting!

A very challenging problem is to estimate the multiplicity of the spectrum (the dimension of the eigenspaces) as the energy levels (the eigenvalues) grow to infinity. The existing results are so poor we do not bother to state anything here.

The next fundamental question about Maass cusp forms is how are the masses distributed. Naturally, they are distributed evenly in an asymptotic sense as the energy level tends to infinity. No excessive accumulation should occur due to the constant negative curvature of the Riemann surface  $\Gamma \backslash \mathbb{H}$  (the habitat in which Maass forms reside). Indeed, Z. Rudnick and P. Sarnak [RS] have a conjecture (known as the Unique Quantum Ergodicity Conjecture) that the measures  $d\mu_j z$  associated with the eigenstates  $u_j(z)$  tend to the Riemannian measure on  $\Gamma \backslash \mathbb{H}$ , precisely

$$d\mu_j z = \frac{|u_j(z)|^2}{\langle u_j, u_j \rangle} d\mu z \rightarrow \frac{d\mu z}{\text{vol}(\Gamma \backslash \mathbb{H})}, \quad \text{as } t_j \rightarrow \infty.$$

It is an important condition that  $u_j(z)$  are primitive Hecke forms, otherwise the assertion is obviously false as the following sequence  $(y^{k/2} f(z))^j$ ,  $j = 1, 2, 3, \dots$  shows, where  $f(z)$  is a fixed classical cusp form of weight  $k$ . Therefore again the arithmeticity plays a role. W. Luo and P. Sarnak [LS] proved the Ergodicity Conjecture on average in the following quantitative format

$$\sum_{0 < t_j \leq T} \left| \int_{\Gamma \backslash \mathbb{H}} f(z) d\mu_j z - \frac{3}{\pi} \int_{\Gamma \backslash \mathbb{H}} f(z) d\mu z \right|^2 \ll T^{1+\varepsilon}.$$

The proof uses heavily analytic properties of the Rankin-Selberg  $L$ -function

$$L(f \otimes f, s) = \sum_1^{\infty} |\lambda_j(n)|^2 n^{-s}.$$

A completely different approach has been recently taken by E. Lindenstrauss. It is based on ergodic methods of G. Margulis, M. Ratner. He did prove the Quantum Ergodicity Conjecture for arithmetic compact surfaces  $\Gamma \backslash \mathbb{H}$  ( $\Gamma$  is a quaternion group), but he barely got the limit with no spare margin. Lindenstrauss arguments [L] appeal to arithmetic from a different angle than the spectral methods and are also used for problems in diophantine approximations.

However, the Quantum Ergodicity Conjecture for the congruence group  $\Gamma = \Gamma_0(q)$  is still open. In this case, the arithmeticity manifests itself beautifully in the Triple Product  $L$ -functions. Suppose for simplicity that  $\Gamma = SL_2(\mathbb{Z})$  and  $u_1, u_2, u_3$  are even Hecke-Maass cusp forms of weight  $k = 0$ . The master  $L$ -functions of these forms are

$$\begin{aligned} L(s, u_j) &= \sum_1^{\infty} \lambda_j(n) n^{-s} \\ &= \prod_p (1 - \alpha_j(p) p^{-s})^{-1} (1 - \beta_j(p) p^{-s})^{-1} \end{aligned}$$

which satisfy the following functional equation

$$\Lambda(s, u_j) = \pi^{-s} \Gamma\left(\frac{s + it_j}{2}\right) \Gamma\left(\frac{s - it_j}{2}\right) L(s, u_j) = \Lambda(1 - s, u_j).$$

The Triple Product  $L$ -function (see [G]) is defined by the following Euler product of degree eight

$$L(s, u_1 \times u_2 \times u_3) = \prod_p \prod_{\gamma_j = \alpha_j, \beta_j} (1 - \gamma_1 \gamma_2 \gamma_3 p^{-s})^{-1}, \text{ if } \operatorname{Re} s > 1.$$

It has analytic continuation to the whole complex plane and it satisfies the functional equation

$$\begin{aligned} \Lambda(s, u_1 \times u_2 \times u_3) &= \pi^{-4s} \prod \Gamma\left(\frac{1}{2}(s \pm it_1 \pm it_2 \pm it_3)\right) L(s, u_1 \times u_2 \times u_3) \\ &= \Lambda(1 - s, u_1 \times u_2 \times u_3). \end{aligned}$$

After fundamental works by Garrett [G] and Harris-Kudla [HK], Thomas Watson [Wa] in a recent thesis gave us a beautiful formula for the central value

$$\Lambda\left(\frac{1}{2}, u_1 \times u_2 \times u_3\right) = \left(\frac{6}{\pi}\right)^2 \Lambda(1, \operatorname{sym}^2 u_1) \Lambda(1, \operatorname{sym}^2 u_2) \Lambda(1, \operatorname{sym}^2 u_3)$$

$$\left| \int_{\Gamma \backslash \mathbb{H}} \frac{u_1(z)u_2(z)u_3(z)}{\|u_1\| \|u_2\| \|u_3\|} d\mu z \right|^2.$$

Here  $\Lambda(s, \text{sym}^2 u_j)$  denotes the completed  $L$ -function for the symmetric square representation attached to  $u_j$ ;

$$\begin{aligned} L(s, \text{sym}^2 u_j) &= \prod_p (1 - \alpha_j^2 p^{-s})^{-1} (1 - \alpha_j \beta_j p^{-s})^{-1} (1 - \beta_j^2 p^{-s})^{-1} \\ \Lambda(s, \text{sym}^2 u_j) &= \pi^{-\frac{3s}{2}} \Gamma\left(\frac{s}{2}\right) \Gamma\left(\frac{s + it_j}{2}\right) \Gamma\left(\frac{s - it_j}{2}\right) L(s, \text{sym}^2 u_j) \\ &= \Lambda(1 - s, \text{sym}^2 u_j). \end{aligned}$$

Note that  $\|u_j\|^2 = \frac{6}{\pi} \Lambda(1, \text{sym}^2 u_j) > 0$ , hence by Watson's formula  $L(\frac{1}{2}, u_1 \times u_2 \times u_3) \geq 0$ , which is consistent with the Riemann hypothesis for  $L(s, u_1 \times u_2 \times u_3)$ . However Watson's formula was created for a more profound consequence, that is towards the Quantum Ergodicity Conjecture. To this end, fix a cusp form  $u_1(z) = u(z)$  and choose  $\bar{u}_2(z) = u_3(z) = u_j(z)$  with varying  $t_j$ . In this case, Watson gets

$$L\left(\frac{1}{2}, u \times u_j \times \bar{u}_j\right) = Q_j \left| \int_{\Gamma \backslash \mathbb{H}} u(z) \frac{|u_j(z)|^2}{\langle u_j, u_j \rangle} d\mu z \right|^2$$

where  $Q_j$  is a product of the involved gamma functions and the values  $L(1, \text{sym}^2 u_j)$ . By Stirling's formula and by

$$(\log t_j)^{-1} \ll L(1, \text{sym}^2 u_j) \ll \log t_j$$

(which is another great result in analytic theory of automorphic  $L$ -functions due to Hoffstein and Lockhart), we find that  $Q_j \asymp t_j^{\frac{3}{2}} (\log t_j)^{\pm 3}$ . Hence, the Quantum Ergodicity Conjecture, which boils down to

$$\int_{\Gamma \backslash \mathbb{H}} u(z) \frac{|u_j(z)|^2}{\langle u_j, u_j \rangle} d\mu z \rightarrow 0, \quad \text{as } t_j \rightarrow \infty,$$

would follow from the bound

$$L\left(\frac{1}{2}, u \times u_j \times \bar{u}_j\right) \ll t_j^{\frac{3}{2} - \delta}$$

with  $\delta$  any positive number, no matter how small. Such bound with  $\delta = 0$  comes straight by the convexity principle for analytic functions (also the functional equation is needed), so we are asking for a subconvexity bound. The Grand Riemann Hypothesis would finish the job, but

getting it done unconditionally is quite a problem. Note that in the special case  $u_1 = u$ ,  $\bar{u}_2 = u_3 = u_j$ , the triple product  $L$ -function factors

$$L(s, u \times u_j \times \bar{u}_j) = L(s, u)L(s, u \otimes \text{sym}^2 u_j),$$

hence we are really faced with a subconvexity problem for  $L$ -functions of degree six (down from eight).

In the last decade, the subject of subconvexity bounds for  $L$ -functions dominated analytic number theory as many interesting problems in arithmetic were reduced to just such a bound. A powerful technology was developed, the amplification method, which produces subconvexity bounds for  $L$ -functions of degree two. The method has been significantly refined and successfully applied to some  $L$ -functions of larger degree by Philippe Michel [M1], see his comprehensive survey [M2] with many beautiful and surprising applications. Other deep results are given by E. Kowalski, P. Michel and J. Vanderkam [KMV].

#### 4. Orthogonality of Hecke Eigenvalues

At first glance it seems obvious that the eigenvalues  $\lambda_f(n)$  are “independent variables” as the cusp form  $f$  varies over a family  $\mathcal{F}$ . Analytic number theory offers an adequate language and tools to quantify such a statement. To cut the exposition, from now on, we consider only the classical holomorphic forms. Because distinct forms in a basis are orthogonal, so seem to be their coefficients. What does it mean? Specifically one should expect that a general vector of complex numbers  $\alpha = (\alpha_n)$ ,  $n \leq N$ , is almost orthogonal to almost all Hecke vectors  $(\lambda_f(n))$ ,  $n \leq N$ . In other words, a general linear form

$$\mathcal{L}_f(\alpha) = \sum_{n \leq N} \alpha_n \lambda_f(n)$$

is quite often small relative to the  $\ell_2$ -norm

$$\|\alpha\|^2 = \sum_{n \leq N} |\alpha_n|^2.$$

Of course, one cannot beat the trivial bound  $|\mathcal{L}_f(\alpha)|^2 \ll N \|\alpha\|^2$  for any given  $f$  (just choose  $\alpha_n = \bar{\lambda}_f(n)$ ), but a non-trivial estimate for the average is possible because distinct  $f \in \mathcal{F}$  pretend they do not see each other. The best estimate one can hope for is

$$\sum_{f \in \mathcal{F}} |\mathcal{L}_f(\alpha)|^2 \ll (|\mathcal{F}| + N) \|\alpha\|^2.$$

Indeed, if  $\mathcal{F}$  is a sufficiently large family, such a result is known in literature as the large sieve inequalities (just a name!). First, the large sieve inequalities were established for the Dirichlet characters ( $GL_1$  forms so to speak) and later in the 1980’s for the  $GL_2$  cusp forms. Needless to say, such kind of estimates came by a demand and they are extremely powerful in applications. In practice, the large sieve inequalities are as good as the Grand Riemann Hypothesis itself (see the classic book of E. Bombieri [B]).

Let me mention a large sieve inequality for the family  $\mathcal{F}_\chi$ , which is an orthonormal basis of  $S_k(\Gamma_0(q), \chi)$ ,  $k \geq 2$ . Define the coefficients  $\psi_f(n)$  by

$$f(z) = \sum_1^\infty \psi_f(n) \left( \frac{(4\pi n)^{k-1}}{\Gamma(k-1)} \right)^{\frac{1}{2}} e(nz)$$

(a new more convenient normalization of Hecke eigenvalues). We have

$$\sum_{f \in \mathcal{F}_\chi} \left| \sum_{n \leq N} \alpha_n \psi_f(n) \right|^2 = \left( 1 + O\left(\frac{N}{q}\right) \right) \|\alpha\|^2$$

where the implied constant depends only on  $k$ . This shows that the normalized coefficients  $\psi_f(n)$  have size about  $q^{-\frac{1}{2}}$  on average; and if  $q \geq N$ , most of the linear forms

$$\mathcal{M}_f(\alpha) = \sum_{n \leq N} \alpha_n \psi_f(n)$$

enjoy a cancellation of terms due to the variation of the argument of  $\psi_f(n)$ . By comparison with the trivial bound, we gained a factor  $N^{\frac{1}{2}}$ , which is the maximal possible.

The larger the family  $\mathcal{F}$  that is put into the play, the stronger orthogonality that comes out. Driven by this principle, X. Li performed (in a forthcoming thesis at Rutgers) extra averaging over the characters  $\chi \pmod{q}$ . Yes, she did get extra savings, but surprisingly not the maximal one which one would hope for by speculating with the Grand Riemann Hypothesis, or by applying generally accepted heuristic arguments. In fact, one of her results asserts that

$$\frac{2}{\varphi(q)} \sum_{\substack{\chi \pmod{q} \\ \chi(-1) = (-1)^k}} \sum_{f \in \mathcal{F}_\chi} |\mathcal{M}_f(\alpha)|^2 = \frac{1}{q} \sum_{h \leq Nq^{-1}} \sum_{\substack{t \leq Nq^{-1} \\ (t,q)=1}} |\mathcal{P}_{ht}(\alpha)|^2 + O(N^\varepsilon \|\alpha\|^2),$$

where  $q$  is prime,  $\alpha_n$  are any complex numbers for  $n \leq N$  with  $q \leq N \leq q^2$  and

$$\mathcal{P}_{ht}(\alpha) = \frac{2\pi}{t} \sum_{n \leq N} \alpha_n S(h\bar{q}, n; t) J_{k-1} \left( \frac{4\pi}{t} \sqrt{\frac{hn}{q}} \right)$$

is the dual linear form. Here,  $S(h\bar{q}, n; t)$  is the original Kloosterman sum and  $J_{k-1}(x)$  is the Bessel function. The error term is superb, but the dual linear form  $\mathcal{P}_{ht}(\alpha)$  can be made exceptionally large. To this end, fix  $h, t$  and choose the coefficients

$$\alpha_n = S(h\bar{q}, n; t) J_{k-1} \left( \frac{4\pi}{t} \sqrt{\frac{hn}{q}} \right)$$

(call them “absolute” Kloosterman sums), showing that

$$\mathcal{P}_{ht}(\alpha) \asymp \left( \frac{qN}{h} \right)^{\frac{1}{4}} \|\alpha\|.$$

Hence, some of the original linear forms  $\mathcal{M}_f(\alpha)$  must be as large as  $\mathcal{P}_{ht}(\alpha)$  above. This is quite a disturbing discovery because it means that the Fourier coefficients  $\psi_f(n)$  (so also the Hecke eigenvalues  $\lambda_f(n)$ ) bend towards the direction of the vector  $\alpha = (\alpha_n)$  of absolute Kloosterman sums. Changing  $h$  and  $t$ , we get a panorama of favorite directions.

A quantitative analysis reveals more surprises, but we have no room to discuss them in these lectures.

### 5. Orthogonality of Elliptic Curves

The demand for orthogonality of Hecke eigenvalues becomes harder when the family is small relative to the size of the conductor. Of special interests are the families of elliptic curves

$$E : y^2 = x^3 + ax + b, \quad \Delta = -16(4a^3 + 27b^2)$$

where  $a, b \in \mathbb{Z}$ ,  $1 \leq a \leq A$ ,  $1 \leq b \leq B$ . The corresponding modular forms  $f_{ab}(z)$  are primitive cusp forms of weight two and level  $q \leq X = 16(4A^3 + 27B^2)$ . We have about  $AB \ll X^{\frac{5}{6}}$  such curves while the total number of all primitive cusp forms of weight two and level  $q \leq X$  is about  $X^2$ . Therefore we are dealing with a very small subset indeed. Moreover the coefficients  $a, b$  in the Weierstrass equation do not need to run over all integers in segments so one can select still smaller subfamilies without destroying their integrity (spectral completeness).

If  $m$  is odd square free, the Hecke eigenvalue for the cusp form  $f_{ab}(z)$  is given by the character sum

$$\lambda_{ab}(m) = \frac{\mu(m)}{\sqrt{m}} \sum_{x \pmod{m}} \left( \frac{x^3 + ax + b}{m} \right)$$

where  $\mu(m)$  is the Möbius function (a vital factor in certain investigations). In addition to  $\lambda_{ab}(m)$ , it is also interesting to consider the reduced character sums

$$\lambda_{ab}^*(m) = \frac{\mu(m)}{\sqrt{m}} \sum_{\substack{x \pmod{m} \\ (x,m)=1}} \left( \frac{x^3 + ax + b}{m} \right).$$

Note that

$$\lambda_{ab}(m) = \sum_{d|m} \frac{\mu(d)}{\sqrt{d}} \left( \frac{b}{a} \right) \lambda_{ab}^* \left( \frac{m}{d} \right).$$

The orthogonality of  $\lambda_{ab}^*(m)$  is very good, as the following estimate shows

Theorem. For any complex numbers  $\alpha_a, \beta_b$  we have

$$\sum_{1 \leq m \leq M} \left| \sum_{1 \leq a \leq A} \sum_{1 \leq b \leq B} \alpha_a \beta_b \lambda_{ab}^*(m) \right| \ll \|\alpha\| \|\beta\| (M + \sqrt{A})(M + \sqrt{B}) M^{-1+\varepsilon}$$

where the implied constant depends only on  $\varepsilon$  (the superscript  $b$  restricts the summation to square free numbers).

Observe that we have stated the result in a fashion which is dual to the large sieve type. By trivial estimation using Hasse's bound  $\lambda_{ab}^*(m) \ll \tau(m)$ , we would only get  $\|\alpha\| \|\beta\| \sqrt{AB}M$ . For applications, one wishes to save slightly more than  $M^{\frac{1}{2}}$ , while our theorem does



satisfy the wish if  $M^2 \ll AB$ . Choosing  $A = X^{\frac{1}{3}}$  and  $B = X^{\frac{1}{2}}$ , we have elliptic curves of discriminant  $\Delta_{ab} = -16(4a^3 + \alpha + b^2) \ll X$  and we obtain the desired saving factor of  $M^{\frac{1}{2}}$  by the theorem, provided  $M \ll X^{\frac{5}{12}}$ . This is quite a large range, yet not completely satisfying. For example, to estimate correctly the Hasse-Weil  $L$ -functions, one has to pass the barrier at  $M = X^{\frac{1}{2}}$ . The following conjecture would open the barrier, and it sounds plausible.

Conjecture. For any complex numbers  $\gamma_m$  supported on square free numbers, we have

$$\sum_{1 \leq a \leq A} \left| \sum_{1 \leq m \leq M} \gamma_m \lambda_{ab}(m) \right|^2 \ll (A + M) \sum_{1 \leq m \leq M} |\gamma_m \tau(m)|^2$$

where the implied constant may depend only on  $b$  slightly.

When  $b$  is a square, we have elliptic curves of positive rank. The conjecture in this case would be helpful for solving the class number problem of the imaginary quadratic fields  $K = Q(\sqrt{D})$ , the goal being

$$h(D) \gg \sqrt{|D|} (\log |D|)^{-1}.$$

The large sieve type inequalities for families of elliptic curves also have applications (indirectly) for estimation of the analytic rank. These questions are addressed with very strong results in recent Ph.D. theses by S. Miller (Princeton) and M. Young (Rutgers).

## References

- [DS-P] W. Duke and R. Schultze-Pillot, Representations of integers by positive ternary quadratic forms and equidistribution of lattice points on ellipsoids, *Invent. Math.* **99** (1990), 49–57.
- [D1] W. Duke, Hyperbolic distribution problems and half-integral weight Maass forms, *Invent. Math.* **92** (1988), 73–90.
- [D2] W. Duke, The dimension of the space of cusp forms of weight one, *IMRN* **2** (1995), 99–109.
- [Se] J-P. Serre, Modular forms of weight one and Galois representations, in Algebraic Number Fields, Academic Press, 1977, 193–268.
- [DS] P. Deligne and J-P. Serre, Formes modulaires de poids 1, *Ann. Sci. Ecole Norm. Sup.* (4) **7** (1974), 507–530.
- [W] A. Wiles, Modular elliptic curves and Fermat’s last theorem, *Ann. of Math.* (2) **141** (1995), 443–551.
- [CI] B. Conrey and H. Iwaniec, The cubic moment of central values of automorphic  $L$ -functions, *Annals of Math.* **156** (2000), 1175–1216.
- [M] H. Maass, Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen, *Math. Ann.* **121** (1969), 141–183.
- [S] A. Selberg, Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series. *J. Indian Math. Soc.* **20** (1958), 47–87.
- [KMV] E. Kowalski, P. Michel, and J. Vanderkam, Rankin-Selberg  $L$ -functions in the level aspect, *Duke Math. J.*
- [G] P. Garrett, Decomposition of Eisenstein series; Rankin triple products, *Annals of Math.* **125** (1987), 209–235.
- [PS] R. Phillips and P. Sarnak, On cusp forms for co-finite subgroups of  $PSL(2, \mathbb{R})$ , *Invent. Math.* **80** (1985), 339–364.
- [RS] Z. Rudnick and P. Sarnak, The behaviour of eigenstates of arithmetic hyperbolic manifolds, *CMP* **161** (1991), 195–213.
- [HK] M. Harris and S. Kudla, The central value of a triple product  $L$ -function, *Annals of Math.* **133** (1991), 605–672.
- [M1] P. Michel, The subconvexity problem for Rankin-Selberg  $L$ -functions with nebentypus and equidistribution of Heegner points, *Annals of Math.* (to appear).
- [M2] P. Michel, Analytic Number Theory and Families of Automorphic  $L$ -functions, *IAS/Park City Mathematics Series* (to appear).
- [B] E. Bombieri, Le Grand Crible dans la Théorie Analytique des Nombres, *Asterisque* **18** (1987).
- [BCDT] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, On the modularity of elliptic curves over  $Q$ , *J. Amer. Math. Soc.* **14** No. 4, (2001), 843–939.
- [Wa] T. Watson, *Rankin triple products and quantum chaos*, Thesis, Princeton University (2001).
- [LS] W. Luo and P. Sarnak, Quantum ergodicity of eigenfunctions on  $SL_2(\mathbb{Z})/H$ , *Publ. Math. IHES* **81** (1995), 207–237.
- [L] E. Lindenstrauss, Invariant measures and arithmetic quantum ergodicity, *Annals of Math.* (to appear).