MODULAR FORMS AND L-FUNCTIONS: A CRASH COURSE

EVAN WARNER

The purpose of this note is to give a quick overview and review of classical modular forms and their L-functions.

1. Modular forms

1.1. The modular curve. Let

$$\mathcal{H} = \{z \in \mathbb{C} : \mathrm{im}(z) > 0\}$$

be the (Poincaré) upper half plane, which we equip with the metric $\frac{dx\,dy}{y^2}$ and an action of $\mathrm{PSL}(2,\mathbb{R})$ given by

$$\gamma z = \frac{az+b}{cz+d}, \qquad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{R}).$$

This is a genuine $\mathrm{PSL}(2,\mathbb{R})$ action as $\pm I$ have the same effect, and it can easily be checked that the action is transitive, isometric, and (for PSL, not SL) faithful. In practice, we will use the $\mathrm{SL}(2,\mathbb{R})$ action because of the advantage in notation given by explicit matrix groups, keeping in mind that this action is "almost faithful" in the above sense.

Consider a discrete subgroup $\Gamma \subset \mathrm{SL}(2,\mathbb{R})$.[1] Mostly, we will be concerned with subgroups of $\mathrm{SL}(2,\mathbb{Z})$, which are automatically discrete, and in fact we will often restrict ourselves to Γ that are not too small in the sense that there exists N such that

$$\Gamma(N) \subset \Gamma \subset \mathrm{SL}(2,\mathbb{R}),$$

where

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) : \begin{pmatrix} z & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \mod N \right\}.$$

Such groups are called *congruence subgroups* and have special arithmetic significance.[2] The integer N is called the *level*.

A *fundamental domain* F for Γ is a connected open subset of $\mathcal{H}$ such that no two points in F are Γ-equivalent and for every point $\tau \in \mathcal{H}$ there is a $g \in \Gamma$ such that $g\tau \in \overline{F}$. For example, if $\Gamma = \Gamma(1) = \mathrm{SL}(2,\mathbb{Z})$, then it can be checked that a fundamental domain for Γ is given by

$$F = \left\{ z \in \mathcal{H} : |z| > 1, |\mathrm{re}(z)| < \frac{1}{2} \right\}.$$

[1] That is, a subgroup that is discrete as a subset with respect to the natural (Euclidean) topology on $\mathrm{SL}(2,\mathbb{R})$. Poincaré proved that a subgroup is discrete if and only if acts discontinuously on $\mathcal{H}$.

[2] Much of the general theory goes through for so-called *Fucshian groups of the first kind*, which in particular include all Γ such that $X(\Gamma)$ has finite volume under the usual (Poincaré) metric, where $X(\Gamma)$ is defined below.

It is useful to note that $\Gamma(1)$ is generated by the elements

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

which often makes the full modular group $\Gamma(1)$ easy to study. Geometrically, T represents a shift to the right by 1, while S represents inversion about the unit circle followed by reflection about the imaginary axis.

We can change variables

$$z \mapsto \frac{z-i}{z+i}$$

to bring $\mathcal{H}$ to the (Poincaré) disc

$$\mathcal{D} = \{z : |z| < 1\}.$$

This change of variables makes evident the symmetry between all the boundary points, which consist of points of the real line $\mathbb{R}$, which map to $\{z : |z| = 1\} \setminus \{1\}$, and the point at infinity, which maps to 1. If we draw the image of a fundamental domain for Γ in $\mathcal{D}$, we see several points at which this image seems to hit the boundary in a narrow point. We would like to characterize these points of the boundary, which are called *cusps*. The correct definition is as follows: a cusp is a point $s \in \mathbb{R} \cup \{\infty\}$ that is fixed by a parabolic element of Γ.[3] Since we will shortly be taking a quotient of $\mathcal{H}$ by Γ, we more commonly let the term *cusp* refer to a Γ-orbit of such points. Thus for the full modular group $\Gamma(1)$ the cusps are precisely $\mathbb{Q} \cup \{\infty\}$ under the former definition, or the singleton set $\{\mathbb{Q} \cup \{\infty\}\}$ under the latter (because $\mathrm{SL}(2,\mathbb{Z})$ is transitive on $\mathbb{Q} \cup \{\infty\}$). We therefore say that $\Gamma(1)$ has one cusp. By restricting to congruence subgroups[4], we ensure that there exist only finitely many cusps.

Let

$$\mathcal{H}^* = \mathcal{H} \cup \{\text{cusps of } \Gamma\},$$

where we take the first definition of cusp, and topologize $\mathcal{H}^*$ as follows: around every point of $\mathcal{H}$, we take the induced (Euclidean) topology. Around each cusp $s \in \mathbb{R}$ we take as a neighborhood basis the set of all sets of the form

$$\{z : |z - (s + ir)| < r\} \cup \{s\}, r \in \mathbb{R}^+.$$

This is the usual open ball with radius r tangent to the real axis at s, *together with the point s itself.* Around the point ∞, if it is a cusp, we take as a neighborhood basis the set of all sets of the form

$$\{z : |\mathrm{re}(z)| > R\}, \quad R \in \mathbb{R}^+.$$

This is *not* the topology induced by the Euclidean topology on $\mathbb{C}$; instead, we have constructed a topology so that the set of all cusps (in the first sense) is discrete. Hausdorffness may be easily checked.

Set

$$X(\Gamma) = \Gamma \setminus \mathcal{H}^*$$

and give $X(\Gamma)$ the quotient topology. We will put charts on $X(\Gamma)$ which will make it into a Riemann surface – the *compactified modular curve*. Around points with "trivial" stabilizer (actually, the stabilizer is $\pm I$), we put ordinary charts, inherited from $\mathcal{H}$. Fortunately, only finitely many non-cusps will not have trivial stabilizer;

[3] Recall that a parabolic element is an element of trace two that is not equal to $\pm I$.

[4] or Fucshian groups of the first kind, more generally.

we call these points *elliptic points*.[5] The stabilizer at an elliptic point is always cyclic. In order to put a legitimate chart around an elliptic point, we have to "mod out" by this stabilizer, which we do by stretching an ordinary chart appropriately. In detail, we map a neighborhood of an elliptic point τ to the origin via

$$z \mapsto \frac{z-\tau}{z-\overline{\tau}},$$

then stretch by $z \mapsto z^n$, where n is the order of the stabilizer of τ, then take the pullback of the ordinary chart around the origin under the composition of the two maps. Around a cusp, which without loss of generality we can consider to be the point ∞, we note that the stabilizer under Γ is isomorphic to $\mathbb{Z}$, generated by $z \mapsto T^n z$ for some n. In order to put a legitimate chart around ∞, we again have to "mod out" by this stabilizer, which we do by mapping $z \mapsto e^{2\pi i z/n}$ and pulling back the usual charts around the origin.[6]

The result is a compact Riemann surface $X(\Gamma)$, the simplest example of a Shimura variety and our basic analytic object.

1.2. **Aside: the modular curve and elliptic curves.** Modular curves for various arithmetically interesting congruence subgroups are *moduli spaces* for – that is, they parametrize in a certain sense – elliptic curves with *level structure*. This is easiest to see in the case of $\Gamma(1)$: an elliptic curve over $\mathbb{C}$ is a complex torus, hence can be described as $\mathbb{C}/\Lambda$ where Λ is a nondegenerate lattice in the complex plane. Two elliptic curves are isomorphic if and only if their corresponding lattices are homothetic. Quotienting out by homotheties can be achieved by bringing each Λ into a standard form (generated by the elements of $\{1, \tau\}$, where $\tau \in \mathcal{H}$), and each nondegenerate lattice has an automorphism group of precisely $\mathrm{SL}(2, \mathbb{Z})$. Therefore we have a bijection

$$\{\text{elliptic curves/isomorphism}\} \leftrightarrow \Gamma(1) \setminus \mathcal{H}.$$

The compactified modular curve $X(\Gamma(1))$ adds at the cusp the "degenerate elliptic curve" with zero discriminant.

This is generalized as follows. Let

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \mod N \right\}$$

and

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma(1) : \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \mod N \right\}$$

which are all congruence subgroups because

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \Gamma(1).$$

Then it is not difficult to show that $X(\Gamma_0(N))$ parametrizes elliptic curves $E/\mathbb{C}$ together with a choice of cyclic subgroup of E of order N, $X(\Gamma_1(N))$ parametrizes elliptic curves $E/\mathbb{C}$ together with a choice of point of E of order N, and $X(\Gamma(N))$ parametrizes elliptic curves $E/\mathbb{C}$ together with a pair of points of order N with a prescribed Weil pairing. For details, see the first chapter of Diamond and Shurman.

[5]In fact, for a suitably generic choice of Γ, there will be no elliptic points at all. For the full modular group $\Gamma(1)$, the elliptic points are i, which has a stabilizer of order 2, and the point $e^{2\pi i/3}$, which has a stabilizer of order 3.

[6]For details of this procedure, see the second chapter of Diamond and Shurman.

More deeply, $X(\Gamma)$ itself will be arithmetically interesting. This is the Eichler-Shimura theory, whereby the Jacobian of $X(\Gamma_0(N))$ can be realized as a scheme over $\mathbb{Z}$ (so if $X(\Gamma_0(N))$ has genus one, it can itself be so realized) and various arithmetic information lines up nicely: there are close relationships between the data of counting points of this scheme over finite fields and the L-function data from modular forms of weight two for $\Gamma_0(N)$.

1.3. **Modular forms.** Now that we have modular curves, the natural next step is to consider holomorphic functions on them. But $X(\Gamma)$, being a compact Riemann surface, has no nonconstant everywhere holomorphic functions via the maximum principle, so we have to extend our inquiry somewhat. Instead, we will be interested in holomorphic functions on $\mathcal{H}$ that, rather than being invariant under Γ, will transform in a particularly simple way.

A *modular form of weight* k *for* Γ, where $k \in \mathbb{Z}$, is a function on $\mathcal{H}$ such that

$$f\left(\frac{az+b}{cz+d}\right) = (cz+d)^k f(z) \qquad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$$

and such that f is holomorphic on $\mathcal{H}$ and at every cusp (that is, there should be no poles anywhere, including "at infinity"). To check holomorphicity at a cusp, which we can assume without loss of generality is the point ∞, we map $z \mapsto e^{2\pi i z/n}$, where n is the smallest integer such that $T^n \in \Gamma$. This maps a neighborhood of ∞ to a punctured disc, and the condition we insist on is that the singularity at the origin is removable; i.e., f extends naturally to each cusp. Note that f does *not* take on a well-defined value at each cusp (because f is not invariant under Γ), but it is well-defined whether the "value" is zero or nonzero.

Let $M_k(\Gamma)$ denote the space of modular forms of weight k for Γ and let $S_k(\Gamma)$ denote the space of *cusp forms* of weight k for Γ; i.e., the subset of $M_k(\Gamma)$ that vanishes at each cusp. Both $M_k(\Gamma)$ and $S_k(\Gamma)$ are naturally vector spaces over $\mathbb{C}$.

For future use, we make one generalization. Let ψ be a Dirichlet character modulo N, and let $M_k(N,\psi)$ be the space of functions, holomorphic everywhere including at every cusp of $\Gamma_0(N)$,[7] such that

$$f\left(\frac{az+b}{cz+d}\right) = \overline{\psi(d)}(ck+d)^k f(z) \qquad \forall \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N).$$

Let $S_k(N,\psi)$ be the set of cusp forms in $M_k(N,\psi)$. These are spaces of *twisted modular forms* and *twisted cusp forms*.[8]

Of course, we have not shown that any modular forms exist for any Γ. To this end, let

$$E_{2k}(z) = \frac{1}{2\zeta(z)} \sum_{(m,n)\neq(0,0)} \frac{1}{(m+nz)^{2k}}$$

be the *Eisenstein series of weight* $2k$. It can be checked that, for $k \geq 2$, $E_{2k} \in M_{2k}(\Gamma(1))$, and as its Fourier expansion is

$$E_{2k}(z) = 1 + \frac{2}{\zeta(1-2k)} \sum_{n=1}^{\infty} \sigma_{2k-1}(n) e^{2\pi i n z}$$

[7](the congruence subgroups $\Gamma_0(N)$ being the most important for our purposes).

[8]There is a useful further generalization involving the concept of a *multiplier system*, which allows us to define modular forms of half-integer weight such as the eta function and the theta function (and its generalizations). For details see sections 2.6 to 2.8 of Iwaniec.

we see that $E_{2k} \notin S_{2k}(\Gamma(1))$. If we let

$$\Delta(z) = \frac{E_4(z)^3 - E_6(z)^2}{1728},$$

then by examining the resulting Fourier series we find that Δ is a cusp form for $\Gamma(1)$ of weight 12.

We have therefore certainly exhibited nontrivial elements of $M_{2k}(\Gamma)$ when $k \geq 2$ and Γ is a congruence subgroup, although in general these spaces will be much larger. For $\Gamma(1)$, however, there is a sense in which we have found all modular forms: for each $k \geq 2$, we have

$$M_{2k}(\Gamma(1)) = S_{2k}(\Gamma(1)) \oplus \mathbb{C} \cdot E_{2k},$$

while the graded ring

$$M(\Gamma(1)) = \bigoplus_{j=0}^{\infty} M_j(\Gamma(1))$$

can be shown to be generated by the Eisenstein series E_4 and E_6, with

$$S(\Gamma(1)) = \bigoplus_{j=0}^{\infty} S_j(\Gamma(1))$$

a principal ideal generated by Δ. It is also not difficult to show that

$$\dim(M_j(\Gamma(1))) = \dim(S_j(\Gamma(1))) + 1 = \begin{cases} \lfloor \frac{k}{12} \rfloor & \text{if } k = 2 \mod 12, \\ \lfloor \frac{k}{12} \rfloor + 1 & \text{otherwise.} \end{cases}$$

While these relations can be derived elementarily, it is rather more difficult to come up with similar dimension formulas for common congruence subgroups. Using the Riemann-Roch theorem and a careful calculation of genus for each modular curve, however, formulas can be derived. As an example, if we let

$$d_N = \frac{1}{2} N^3 \prod_{p|N} \left(1 - \frac{1}{p^2}\right),$$

then if $N \geq 3$ and $k \geq 3$ we have

$$\dim(M_k(\Gamma(N))) = \frac{(k-1)d_N}{12} + \frac{d_N}{2N},$$
$$\dim(S_k(\Gamma(N))) = \frac{(k-1)d_N}{12} - \frac{d_N}{2N},$$

while if $N \geq 5$ and $k \geq 3$ we have

$$\dim(M_k(\Gamma_1(N))) = \frac{(k-1)d_N}{12N} + \frac{1}{4} \sum_{d|N} \phi(d)\phi\left(\frac{N}{d}\right),$$
$$\dim(S_k(\Gamma_1(N))) = \frac{(k-1)d_N}{12N} - \frac{1}{4} \sum_{d|N} \phi(d)\phi\left(\frac{N}{d}\right).$$

Formulas for $\Gamma_0(N)$ is are even more complicated, and formulas for small values of k and N must be computed separately. For details, see the third chapter of Diamond and Shurman.

Having constructed the vector space $S_k(\Gamma)$, we now proceed to turn it into an inner product space by defining the *Petersson inner product*, as follows: if $f, g \in S_k(\Gamma)$, we define

$$(f, g) = \iint_F f(z)\overline{g(z)}y^k \, \frac{dx\, dy}{y^2},$$

where F is any fundamental domain. This is independent of the choice of F because both $f(z)\overline{g(z)}y^k$ and $\frac{dx\,dy}{y^2}$ are invariant under Γ. The double integral converges due to the following lemma and the fact that Γ, being restricted to be a congruence subgroup (or more generally a Fucshian group of the first kind), has a fundamental domain of finite volume:

Lemma 1.1. *Let $f \in M_k(\Gamma)$. Then $f \in S_k(\Gamma)$ if and only if there exists a constant M such that*

$$y^{k/2}|f(x + iy)| \leq M.$$

Proof. It is clear that $g(z) = (\text{im}(z))^{k/2} f(z)$ is Γ-invariant. Without loss of generality, we can bring each cusp to ∞ and therefore consider only the cusp at ∞. Certainly $|g(z)| \leq M$ implies that f vanishes at ∞. In the other direction, if f vanishes at ∞ then g is continuous on $X(\Gamma)$, a compact space, hence g is bounded. □

In fact, as one can check by examining Fourier expansions, cusp forms must decrease rapidly at all cusps.

The space of twisted cusp forms $S_k(N, \psi)$ is also seen to become an inner product space with the Petersson inner product.

1.4. **Hecke operators.** The Hecke operators, once we describe them, will form a commuting family of normal operators on the inner product space $S_k(N, \psi)$, and hence (by the spectral theorem) will possess an orthonormal basis of eigenfunctions. In the case where ψ is trivial, the operators will be self-adjoint, and hence the eigenvalues will all be real. Their importance lies in their arithmetic significance: the eigenvalues we get will turn out to be the Fourier coefficients of the corresponding (normalized) eigenfunction, which together give an L-function (so in particular, the Fourier coefficients of a simultaneous eigenfunction for all the Hecke operators will be multiplicative, as we will describe below).

We will start with the most general description that we will need: Hecke operators on $\Gamma_0(N)$ twisted by a Dirichlet character ψ.[9] Fixing a weight k, let $\mathrm{GL}^+(2, \mathbb{R})$ act on functions on $\mathcal{H}$ by

$$f|_{\gamma,\psi}(z) = \overline{\psi(d)}(\det \gamma)^{k/2}(cz+d)^{-k} f\left(\frac{az + b}{cz + d}\right), \quad \text{where } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}^+(2, \mathbb{R}).$$

Let Σ be the set of all primes dividing N and, denoting the localization of $\mathbb{Z}$ at Σ by $\mathbb{Z}_\Sigma$, define $G_0(N)$ to be the subgroup of $\mathrm{GL}^+(2, \mathbb{Z}_\Sigma)$ such that the lower left

[9]There is a sense in which this situation (i.e., $\Gamma_0(N)$ with twists) encompasses all congruent subgroups: It can be shown that any congruence modular form can be brought into $M_k(\Gamma_1(N))$ for some N by the $\mathrm{GL}^+(2, \mathbb{R})$ action, and we also have that

$$S_k(\Gamma_1(N)) = \bigoplus_\psi S_k(\Gamma_0(N), \psi),$$

where the sum is over all Dirichlet characters modulo N and orthogonal with respect to the Petersson inner product.

hand entry is contained in $N\mathbb{Z}_\Sigma$. In particular, for the full modular group, $G_0(N)$ is just $\mathrm{GL}^+(2,\mathbb{Z})$. Consider the set of double cosets

$$R_N = \Gamma_0(N) \setminus G_0(N)/\Gamma_0(N),$$

which we will call the *Hecke operators*.

We will define an action of R_N on $S_k(\Gamma_0(N), \psi)$ as follows. Abstractly, a double coset can be decomposed into a union of right cosets, and we define $\alpha_i \in \mathrm{GL}^+(2,\mathbb{R})$ by

$$\Gamma_0(N)\alpha\Gamma_0(N) = \bigcup_i \Gamma_0(N)\alpha_i$$

for each $\alpha \in \mathrm{GL}^+(2,\mathbb{R})$. We then let the set R_N act on $S_k(\Gamma_0(N), \psi)$ by

$$T_\alpha(f) = \sum_i f_{\alpha_i,\psi}.$$

It can be checked that for each $\alpha, \beta \in \mathrm{GL}^+(2,\mathbb{R})$, there is a unique $\gamma \in \mathrm{GL}^+(2,\mathbb{R})$ such that

$$T_\beta(T_\alpha(f)) = T_\gamma(f)$$

and that T_α depends only on the double coset $\Gamma_0(N)\alpha\Gamma_0(N)$. Therefore R_N assumes a ring structure, and we have defined an action of R_N on the vector space $S_k(\Gamma_0(N), \psi)$.

The two main facts about R_N are the following: first, R_N is commutative, and second,

$$(T_\alpha(f), g) = \psi(d)(f, T_\alpha(g))$$

for every $f, g \in S_k(\Gamma_0(N), \psi)$. The first fact is somewhat nontrivial; I will only note here that it is proved by first noting that $R_N \subset R_1$ and then demonstrating the assertion in the simpler case of the full modular group, where it can be proven by explicit computation. The second fact, however, is straightforward, and immediately shows that each T_α is normal. As promised, therefore, the spectral theorem for normal operators shows that there exists an orthonormal basis for $S_k(\Gamma_0(N), \psi)$ consisting of simultaneous eigenvalues for all the Hecke operators.

More concretely, still at the level of generality of $\Gamma = \Gamma_0(N)$ with a twist ψ, the Hecke operators consist of the operators $T(n)$ with $(n, N) = 1$, defined by

$$f_{T(n)}(z) = \frac{1}{n} \sum_{ad=n} \psi(a) a^k \sum_{0 \le b < d} f\left(\frac{az+b}{d}\right).$$

(we have adopted Iwaniec's normalization, which is slightly if unimportantly different than our first abstract definition). I will mention only in passing that these operations are more natural than they seem: if we view modular functions in the setting of lattices in the complex plane, then the Hecke operator $T(n)$ coincides with a sum over all sublattices of index n.

2. L-functions

2.1. **Generalities.** Loosely following Selberg, we will "define" an L-function as a Dirichlet series

$$L(s) = \sum_{n=1}^{\infty} \frac{a(n)}{n^s},$$

where $a(n)$ is a sequence of complex numbers, satisfying the following four conditions:

(1) We have a polynomial bound on the growth of $a(n)$, so $L(s)$ converges in some half-plane.[10]
(2) $L(s)$ has an analytic, or at least meromorphic, continuation to a function on $\mathbb{C}$.[11]
(3) $L(s)$ possesses a *functional equation* of some sort relating values on the right half-plane to values on the left half-plane.
(4) $L(s)$ possesses an *Euler product*, which corresponds formally to the multiplicativity of the coefficients $a(n)$.

2.2. **Example: Dirichlet L-functions.** In order to illustrate this definition, we introduce a simple example. Take a Dirichlet character $\chi : \mathbb{Z} \to S^1$ and define

$$L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}.$$

Condition (1) is obvious. The Euler product, condition (4), is trivial to establish formally and only slightly more difficult to prove rigorously; we have

$$L(s, \chi) = \prod_p \frac{1}{1 - \chi(p)p^{-s}},$$

the product being taken over all primes. Conditions (2) and (3) follow from the following:

Theorem 2.1. *Let χ have conductor N, $\epsilon \in \{0, 1\}$ be defined such that $\chi(-1) = (-1)^\epsilon$, and define the* completed L-function

$$\Lambda(s, \chi) = \pi^{-(s+\epsilon)/2}\Gamma\left(\frac{s+\epsilon}{2}\right) L(s, \chi).$$

Then $\Lambda(s, \chi)$ is an entire function of s, except for poles at $s = 0$ and $s = 1$ if and only if χ is trivial, and we have the functional equation

$$\Lambda(s, \chi) = (-i)^\epsilon \tau(\chi) N^{-s} \Lambda(1 - s, \overline{\chi}).$$

Proof sketch. We have two cases, depending on the value of ϵ. First assume $\chi(-1) = 1$, and set

$$\theta_\chi(t) = \frac{1}{2} \sum_{n=-\infty}^{\infty} \chi(n) e^{-\pi n^2 t}.$$

This is the theta function associated with χ. By a suitably generalized ("twisted") version of Poisson's summation formula, a quick calculation of the Fourier transform of θ_χ, and analytic continuation, we arrive at the formula

$$\theta_\chi(t) = \frac{\tau(\chi)}{N\sqrt{t}} \theta_{\overline{\chi}}\left(\frac{1}{N^2 t}\right)$$

(this being the "inversion formula," originally due to Jacobi). Then take the Mellin transform

$$\theta_\chi(t) \mapsto \int_0^\infty \theta_\chi(t) t^{s/2} \frac{dt}{t}$$

of both sides and unravel to get the result.

[10]In practice, we can renormalize and assume a bound of $O(n^\epsilon)$ for all $\epsilon > 0$, which is more natural in many respects.

[11]We should also require that this function has finite order, and that we have some sort of tight control over the poles, if they exist.

In the case where $\chi(-1) = -1$, we have to select a slightly different theta function, for otherwise our sum cancels in pairs and we end up with the zero function. Therefore in this case let

$$\theta_\chi(t) = \frac{1}{2}\sum_{n=-\infty}^{\infty} n\chi(n)e^{-\pi n^2 t},$$

prove a similar inversion formula, take the Mellin transform

$$\theta_\chi(t) \mapsto \int_0^\infty \theta_\chi(t)t^{(s+1)/2}\frac{dt}{t}$$

and unravel to get the result. □

This method of proof – selecting an appropriate theta function, proving an inversion formula for it, and then taking a Mellin transform of the whole business – was originally developed by Riemann to prove the functional equation for the zeta function (taking χ to be the trivial character in the above theorem). It can be generalized to Hecke's proof of the functional equation for L-functions attached to number fields, although the details are formidable.

2.3. **L-functions from modular forms.** The examples above, including Hecke's generalizations, belong to the theory of automorphic forms over GL(1).[12] The first examples of automorphic forms over GL(2) come from the classical modular forms in the following way.

Let $f \in S_k(\Gamma_0(N), \psi)$. Because

$$T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \Gamma_0(N),$$

the function f is periodic of period one, so it has a Fourier expansion[13]

$$f(z) = \sum_{n=1}^{\infty} a(n)e^{2\pi i n z}.$$

Let

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

Since we will also want to consider twists, take a Dirichlet character χ of conductor r such that $(r, N) = 1$, and let

$$L(s, f, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)a(n)}{n^s}.$$

We can think of $L(s, f, \chi)$, at least formally, as the Mellin transform of f twisted by χ.

Let's examine our L-function criteria. Lemma 1.1 implies immediately that we have the bound

$$a(n) = O(n^{k/2}),$$

[12]Meaning, in part, that ultimately they are "explained" by the representation theory of the circle group.

[13]In general, there is a Fourier expansion at each cusp, and there is no need to limit ourselves to $\Gamma_0(N)$ to define them, but a scaling factor may be required. For ease of exposition we will only consider the Fourier expansion at the cusp at infinity for the group $\Gamma_0(N)$.

settling (1).[14]

Criteria (2), (3), and (4) are dealt with by the following theorem, which shows that $L(s, f)$ is a full-fledged L-function when f is a simultaneous eigenfunction of the Hecke operators.

Theorem 2.2 (Hecke)**.** *Assume $f \in S_k(\Gamma_0(N), \psi)$, where ψ is a Dirichlet character modulo N. Let*

$$\Lambda(s, f, \chi) = (r^2 N)^{s/2}(2\pi)^{-s}\Gamma(s)L(s, f, \chi).$$

Then Λ has an analytic continuation to all of $\mathbb{C}$, bounded on vertical strips, and satisfies the functional equation

$$\Lambda(s, f, \chi) = i^k \psi(r)\chi(N)\frac{\tau(\chi)^2}{r}\Lambda(k - s, \tilde{f}, \overline{\chi}),$$

where $\tau(\chi)$ is the usual Gauss sum and

$$\tilde{f}(z) = \frac{1}{N^{k/2}z^k} f\left(-\frac{1}{Nz}\right).$$

If $f \neq 0$ is normalized so that $a(1) = 1$, then f is a simultaneous eigenfunction of all the Hecke operators if and only if $L(s, f)$ has the Euler product

$$L(s, f) = \prod_p \frac{1}{1 - a(p)p^{-s} + \psi(p)p^{k-1-2s}}.$$

Furthermore, $a(p)$ is precisely the eigenvalue of the action of $T(p)$ on f whenever $(p, N) = 1$.

Proof sketch. The analytic continuation and functional equation, though complicated to state, are not difficult to prove; they follow from the modularity of f and then by taking a Mellin transform (note the analogy to Dirichlet L-functions: we get a functional equation from a simpler object first, like a theta function or a modular form, and then take a Mellin transform).

We will sketch a proof of the existence of the Euler product for $N = 1$ and ψ trivial, for ease of exposition. The idea is simply to calculate what happens to the Fourier coefficients under the action of a Hecke operator. In our case, if we let

$$f|_{T(p)}(z) = \sum_{n=1}^{\infty} b(n)e^{2\pi i n z},$$

then it is a simple calculation that

$$b(n) = \sum_{ad=p} \left(\frac{a}{d}\right)^{k/2} \cdot d \cdot a\left(\frac{nd}{a}\right).$$

From this, it is easy to see that if f is an eigenform for all the Hecke operators with eigenvalues c_p, then $c_p = a(p)$ for all p,

$$a(n)a(m) = a(nm) \tag{1}$$

[14]For cusp forms that are simultaneous eigenforms for the Hecke operators, we can do better. With a great deal of effort, one can prove that in that case

$$a(n) = O(n^{(k-1)/2+\epsilon})$$

for every $\epsilon > 0$. This is the famous Ramanujan-Petersson conjecture, which was settled by Deligne in the 1970s using methods of algebraic geometry.

whenever $(n, m) = 1$, and

$$a\left(p^{r+1}\right) - a(p)a\left(p^r\right) + p^{k-1}\left(p^{r-1}\right) = 0. \tag{2}$$

From (1), we get

$$L(s, f) = \prod_p \left(\sum_{r=0}^{\infty} a\left(p^r\right) p^{-rs}\right),$$

at least formally, and then from (2) we get the desired Euler product formula. This argument is reversible. □

2.4. **Modular forms from L-functions.** There are a few questions left unanswered from Hecke's theory relating to "going backwards;" i.e., determining a modular form from its L-function. The first is the following: if $f_1, f_2 \in S_k(\Gamma_0(N), \psi)$ have the same eigenvalues for all of the Hecke operators (that is, all of the $T(p)$ for $p \nmid N$), are they multiples? In other words, do the simultaneous eigenspaces of the $T(p)$ all have dimension one?

The short answer to this question is no. For example, $S_{12}(\Gamma_0(2))$ contains both the functions $\Delta(z)$ and $\Delta(2z)$, which are easily seen to be linearly independent but share the same eigenvalues (up to multiplicity). The problem occurs because $\Delta(z)$ does not "genuinely belong" at level two; it is a holdover from $S_{12}(\Gamma_0(1))$. However, this is the only kind of problem that can occur: if we only consider functions that "belong at the right level," then we have the following multiplicity-one theorem:

Theorem 2.3 (Artin-Lehner). *Let the linear subspace of $S_k(\Gamma_0(N))$ consisting of modular forms of the form $g(dz)$, where $d|N$, $d \neq 1$, and $g(z) \in S_k(\Gamma_0(N/d))$, be called the space of* oldforms *in $S_k(\Gamma_0(N))$. Let the space of* newforms *be its orthogonal complement. Then $S_k(\Gamma_0(N))$ has a basis of Hecke eigenforms consisting of oldforms and newforms, and two newforms with the same eigenvalues for all of the Hecke operators are multiples of one another.*[15] □

The second question we will raise is the question of *converse theorems*: given an L-function, when does it come from a modular form? That is, if we have an L-function

$$L(s) = \sum_{n=1}^{\infty} \frac{a(n)}{n^s} \tag{3}$$

and we define

$$f(z) = \sum_{n=1}^{\infty} a(n) e^{2\pi i n z},$$

when does f lie in some space $S_k(\Gamma_0(N), \psi)$? Hecke proved a theorem to the effect that if the $a(n)$ are polynomially bounded, the L-function has an analytic equation (bounded in vertical strips) and functional equation as expected, then f will be in $S_k(\Gamma(1))$, and if the L-function has an Euler product of the right shape then f will be a simultaneous eigenform for the Hecke operators.

The generalization to higher level, due to Weil, is more complicated. Our definition of $f(z)$ ensures that f will be periodic of period one, hence modular with respect to the element $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. In the level one case, the functional equation

[15] In fact, this requirement can be relaxed to "all but finitely many Hecke operators."

yields modularity with respect to the matrix $S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and because S and T generate $\Gamma(1)$, we're done. In general, however, $\Gamma_0(N)$ will have many more generators, so we have to assume more to start. The solution is that if we assume a functional equation for our L-function *and for many twists of our L-function*, we can recover modularity of f. Specifically, we need to consider all twists by characters of conductor r, where $(r, N) = 1$.

Weil's converse theorem (and its generalizations to other automorphic forms) are not merely curiosities; they are essential in further study along the lines of the Langlands program.

3. References

Bump, D., *Automorphic Forms and Representations*, Cambridge Studies in Advanced Mathematics, Vol. 55, 1998.

Diamond, F. and J. Shurman, *A First Course in Modular Forms*, Graduate Texts in Mathematics, Vol. 228, 2000.

Garrett, P., *Analytic continuation, functional equation: examples*, notes accessed at `http://www.math.umn.edu/~garrett/`, 2012.

Gelbart, S., *Automorphic Forms on Adele Groups*, Annals of Mathematics Studies, Vol. 83, 1975.

Iwaniec, H., *Topics in Classical Automorphic Forms*, Graduate Studies in Mathematics, Vol. 17, 1997.