Lecture Notes: An invitation to modular forms

October 3, 2011

1 Tate's thesis in the context of class field theory

1.1 Reciprocity laws

Let K be a number field or function field. Let $C_K = \mathbf{A}_K^\times / K^\times$ be the idele class group. One of the main theorems of class field theory is the existence of a *reciprocity map*

$$\mathrm{rec}_K : C_K \to \mathrm{Gal}(K^{\mathrm{ab}}/K).$$

This map is continuous with dense image. Whenever L/K is a finite abelian extension, there is a relative version of the above map, call it $\mathrm{rec}_{L/K}$, that makes the diagram commute:

$$\begin{array}{ccc} C_K & \xrightarrow{\mathrm{rec}_K} & \mathrm{Gal}(K^{\mathrm{ab}}/K) \\ \downarrow & & \downarrow \\ C_K/N_{L/K}(C_L) & \xrightarrow[\mathrm{rec}_{L/K}]{\sim} & \mathrm{Gal}(L/K) \end{array}$$

(There is also a diagram which relates rec_K to rec_L.)

For each place v of K there is a *local reciprocity map*:

$$\mathrm{rec}_{K_v} : K_v^\times \to \mathrm{Gal}(K_v^{\mathrm{ab}}/K_v).$$

There is a relative version of rec_{K_v} as well. In addition, when v is a nonarchimedean place, we recognize that either side of the above has a distinguished quotient: $\mathbf{Z}$ on the left side, and $\mathrm{Gal}(k_v^{\mathrm{ab}}/k_v) \cong \hat{\mathbf{Z}}$ on the right. The local reciprocity map has the property that the diagram

$$\begin{array}{ccc} K_v^\times & \xrightarrow{\mathrm{rec}_{K_v}} & \mathrm{Gal}(K_v^{\mathrm{ab}}/K_v) \\ {\scriptstyle v}\downarrow & & \downarrow \\ \mathbf{Z} & \longrightarrow & \hat{\mathbf{Z}} \end{array}$$

commutes. Here we are identifying $\hat{\mathbf{Z}}$ with $\mathrm{Gal}(k_v^{\mathrm{ab}}/k_v)$ by matching up $1 \in \hat{\mathbf{Z}}$ to the arithmetic Frobenius (the automorphism sending x to x^q, where $q = \#\mathcal{O}_v/\mathfrak{p}_v$). In other words, rec_{K_v} sends uniformizers to arithmetic Frobenius elements. (There is another, and perfectly good, normalization of local class field theory, where uniformizers get sent to geometric Frobenius elements, in which case the definitions of the recs have to be inverted.)

The reciprocity maps exhibit *local-global compatibility*, meaning that

$$\begin{array}{ccc} K_v^\times & \xrightarrow{\mathrm{rec}_{K_v}} & \mathrm{Gal}(K_v^{\mathrm{ab}}/K_v) \\ \downarrow & & \downarrow \\ C_K & \xrightarrow{\mathrm{rec}_K} & \mathrm{Gal}(K^{\mathrm{ab}}/K) \end{array}$$

commutes.

1.2 Artin L-functions and Hecke L-functions

Consider the consequences for Artin L-functions. Let K be a number field and let $\sigma\colon \mathrm{Gal}(K^{\mathrm{ab}}/K) \to \mathbf{C}^\times$ be a (necessarily finite-order) one-dimensional Artin character. The associated Artin L-function is

$$L(\sigma, s) = \prod_{\mathfrak{p}} \left(1 - \frac{\sigma(\mathrm{Frob}_{\mathfrak{p}})}{N\mathfrak{p}^s}\right)^{-1},$$

where the product runs over the primes of K at which σ is unramified (meaning that σ is trivial on the inertia group of $\mathfrak{p}$).

Let $\chi = \sigma \circ \mathrm{rec}$; this is a Hecke character. Then χ and σ are unramified at the same places. For an unramified prime $\mathfrak{p}$ with uniformizer $\pi_{\mathfrak{p}}$, we have $\chi(\pi_{\mathfrak{p}}) = \sigma(\mathrm{Frob}_{\mathfrak{p}})$ (because of local-global compatibility, and because $\mathrm{rec}_{K_{\mathfrak{p}}}$ takes uniformizers to arithmetic Frobenius elements. Therefore the Artin L-function is a Hecke L-function:

$$L(\sigma, s) = L(\chi, s)$$

As a consequence, $L(\sigma, s)$ admits an analytic continuation to all $s \in \mathbf{C}$ (which is even entire if σ is nontrivial). If $\Lambda(\chi, s)$ is the completed L-function (throw in the Γ-factors for each infinite place), then $\Lambda(\chi, s)$ is related to $\Lambda(\chi^{-1}, 1-s)$ via the functional equation:

$$L(\chi^{-1}, 1-s) = (\text{fudge}) \prod_{v \in S} \rho_v(\chi, s) L(\chi, s)$$

Here "fudge" only depends on the ramified primes in $K/\mathbf{Q}$ and essentially not on χ. The local factor $\rho_v(\chi, s)$ only depends on the restriction χ_v of χ to K_v. (This is the innovation which pushed Tate's results beyond Hecke's.) Therefore $\rho_v(s)$ only depends on the restriction of σ to the decomposition group at v.

What sorts of (analytic) objects π could be associated to n-dimensional Artin L-functions? Whatever they are, they should have:

1. An L function $L(\pi, s)$ which is a meromorphic function of all $s \in \mathbf{C}$, such that $L(\pi, s) = L(\sigma, s)$ for s with large enough real part:

2. Local components π_v, which only depend on the restriction of σ to the decomposition group at v,

3. For each v, a local factor $\rho(\pi_v, s)$, which only depends on π_v, such that $L(\pi, s)$ has a functional equation of the form

$$L(\check{\pi}, k - s) = \prod_v \rho(\pi_v, s) L(\pi, s)$$

   up to a fudge factor.

2 Modular Forms: Definitions

For $K = \mathbf{Q}$ and $n = 2$, the answer turns out to be modular forms and their cousins the Maass forms, although we will have to put in a significant amount of work to see why these objects have the desired properties. (A modular form is a function on the upper half-plane. What could it mean to take its local component at 17?)

The classical theory of modular forms is due to Hecke. It has all to do with the action of discrete subgroups of $\mathrm{SL}_2(\mathbf{R})$ (known as Fuchsian groups) on the upper half plane $\mathcal{H}$. We start with the *principal congruence subgroup*

$$\Gamma(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z}) \;\middle|\; a, d \equiv 1 \pmod N,\ b, c \equiv 0 \pmod N \right\}$$

and say that a subgroup $\Gamma \subset \mathrm{SL}_2(\mathbf{Z})$ is a *congruence subgroup* if it contains $\Gamma(N)$ for some N. Two important examples are $\Gamma_0(N)$ (resp., $\Gamma_1(N)$), which are those subgroups of $\mathrm{SL}_2(\mathbf{Z})$ where $c \equiv 0 \pmod N$ (resp., $c \equiv 0 \pmod N$ and $a \equiv b \equiv 1 \pmod N$).

Let Γ be a congruence subgroup, and let k be an integer. A *modular form* of weight k for Γ is a complex-valued function f on the upper half-plane $\mathcal{H}$ with the following properties:

1. For all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma$,

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

2. f is holomorphic on $\mathcal{H}$.

3. f is holomorphic at the cusps of $\Gamma \backslash \mathcal{H}$.

For any $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbf{R})$ with positive determinant, we write

$$f|_{g,k}(z) = (cz + d)^{-k} (\det g)^{k/2} f(gz);$$

this actually defines an action of $\mathrm{GL}_2^+(\mathbf{R})$ on functions on $\mathcal{H}$. (The center acts trivially, so that in fact we have an action of $\mathrm{PGL}_2^+(\mathbf{R})$.) The first condition can be restated as $f|_{\gamma,k} = f$ for all $\gamma \in \Gamma$.

The last point requires some explanation. A *cusp* of Γ is an equivalence class in an element of $\mathbf{R} \cup \{\infty\}$ which is fixed by a parabolic subgroup of Γ (a subgroup which has exactly one fixed point on the Riemann sphere). For instance, ∞ is always fixed by the intersection of $\begin{pmatrix} 1 & \mathbf{Z} \\ & 1 \end{pmatrix}$ with Γ, so ∞ is always a cusp of a congruence subgroup. We now define what it means for f to be holomorphic at ∞. The subgroup $\Gamma_\infty \subset \Gamma$ which fixes ∞ is of the form $\begin{pmatrix} 1 & \mathbf{Z}w \\ & 1 \end{pmatrix}$ for some $w \geq 1$, known as the width of the cusp at ∞. We have $f(z+w) = f(z)$ for $z \in \mathcal{H}$, so that it makes sense to define $\hat{f}(q)$ on the domain $0 < |q| < 1$ by

$$\hat{f}(e^{2\pi i z/w}) = f(z)$$

Then the condition of holomorphicity at ∞ is the condition that $\hat{f}$ extend to a holomorphic function on $|q| < 1$. For general cusps c, one can always find a $\gamma \in \mathrm{SL}_2(\mathbf{Z})$ which translates ∞ onto c. Then the condition that f be holomorphic at c is understood to be the same as the condition that $f|_{\gamma,k}$ (which is attempting to be a modular form for $\gamma^{-1}\Gamma\gamma$) be holomorphic at ∞. One must check that this definition does not depend on the choice of γ.

We very often consider the Taylor series expansion of $\hat{f}(q)$ around $q = 0$, which converges for all $z \in \mathcal{H}$:

$$f(z) = \sum_{n \geq 0} a_n e^{2\pi i n z/w}.$$

Note that in the cases of $\Gamma_0(N)$ and $\Gamma_1(N)$, the width of ∞ is $w = 1$.

Write $M_k(\Gamma)$ for the complex vector space of modular forms of weight k on Γ. Write $S_k(\Gamma)$ for the space of modular forms which are *cusp forms*, which means they are 0 at every cusp of Γ. It turns out that $M_k(\Gamma)$ is a finite-dimensional vector space, so that $S_k(\Gamma)$ is a finite-dimensional Hilbert space.

There is a bilinear pairing $S_k(\Gamma) \times S_k(\Gamma) \to \mathbf{C}$ known as the Petersson inner product:

$$(f,g)_{k,\Gamma} = \int_{\mathcal{F}} f(z)\overline{g(z)} y^k \frac{dxdy}{y^2},$$

where $\mathcal{F}$ is a fundamental domain for $\Gamma\backslash\mathcal{H}$. (Check that the integrand really is Γ-invariant!)

It is generally far easier to produce examples of elements of $M_k(\Gamma)$ than it is to produce cusp forms in $S_k(\Gamma)$. For k even, we have the *Eisenstein series*

$$E_k(z) = \frac{1}{2\zeta(k)} \sum_{(c,d)} \frac{1}{(cz+d)^k},$$

where the sum is over all pairs of integers $(c,d) \neq (0,0)$. E_k is very easily seen to ba modular form for $\mathrm{SL}_2(\mathbf{Z})$ of weight k. A little manipulation shows that its Fourier expansion is

$$E_k(z) = 1 + \frac{2}{\zeta(1-k)} \sum_{n \geq 1} \sigma_{k-1}(n) q^n, \; q = e^{2\pi i z}.$$

In can be shown that the graded ring $\oplus_{k \text{ even}} M_k(\mathrm{SL}_2(\mathbf{Z}))$ is generated over $\mathbf{C}$ by the elements E_4 and E_6. As a result $M_{12}(\mathrm{SL}_2(\mathbf{Z}))$ is spanned by E_4^3 and E_6^2, and

$$\Delta = \frac{1}{1728}(E_4^3 - E_6^2) \in S_{12}(\mathrm{SL}_2(\mathbf{Z}))$$

spans the (one-dimensional) space of cusp forms for $\mathrm{SL}_2(\mathbf{Z})$.

The Dirichlet series attached to a modular form $f = \sum_{n\geq 0} a_n q^n$ on $\Gamma_0(N)$ is

$$L(f,s) = \sum_{n\geq 1} \frac{a_n}{n^s}$$

This is related to the Mellin transform of f:

$$(2\pi)^{-s}\Gamma(s)L(f,s) = \int_0^\infty (f(iy) - a_0)y^{s-1}dy$$

Now for the Eisenstein series E_k we have (up to a constant)

$$L(E_k,s) = \sum_{n\geq 1} \frac{\sigma_{k-1}(n)}{n^s} = \zeta(s)\zeta(s-k+1),$$

and this function extends via analytic continuation to all complex s, with a functional equation relating $L(E_k,s)$ to $L(E_k,k-s)$. More provocatively, we can consider the Galois representation

$$\begin{aligned} \rho_\ell \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) &\to \mathrm{GL}_2(\mathbf{Q}_\ell) \\ \sigma &\mapsto \begin{pmatrix} \chi_{\mathrm{cycl}}^{k-1} & \\ & 1 \end{pmatrix}, \end{aligned}$$

where $\chi_{\mathrm{cycl}} \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathbf{Z}_\ell^\times$ is the ℓ-adic cyclotomic character, defined by the condition $\zeta^\sigma = \zeta^{\chi_{\mathrm{cycl}}(\sigma)}$ whenever ζ is an ℓ-power root of 1. (We have $\chi_{\mathrm{cycl}}(\mathrm{Frob}_p) = p$ whenever $p \neq \ell$.) Then ρ_ℓ isn't an Artin representation (it isn't finite-order). Still, it seems we can go ahead and define an L-function anyway. Seeing as the expression $P_p(T) = \det \rho_k(1 - T\cdot \mathrm{Frob}_p)$ always lies in $\mathbf{Z}[T]$ for $p \neq \ell$, with the result not depending on ℓ, the L-function of ρ_ℓ is

$$L(\rho_\ell, s) = \prod_p P_p(p^{-s})^{-1} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}\left(1 - \frac{p^{k-1}}{p^s}\right)^{-1} = L(E_k,s)$$

(again up to a constant).

Much, much less obvious is the existence (due to Deligne) of a family of ℓ-adic Galois representations $\rho_\ell \colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_2(\mathbf{Q}_\ell)$ which has the same properties with respect to cusp forms such as Δ.

3 Hecke's theory, part I: the case of level 1

3.1 Formalism in terms of double cosets

Suppose Γ and Γ' are two congruence subgroups, and $g \in \mathrm{GL}_2^+(\mathbf{Q})$. Then the double coset space $\Gamma g \Gamma'$ can be written as a disjoint union of finitely many left cosets:

$$\Gamma g \Gamma' = \coprod_j \Gamma g_i,\ g_i \in \mathrm{GL}_2^+(\mathbf{Q})$$

If f belongs to $M_k(\Gamma)$, then let

$$f|_{\Gamma g \Gamma'} = \sum_i f|_{g_i}.$$

(Here we are writing $f|_{g_i}$ for $f|_{g_i,k}$, because the weight k is fixed in this discussion.) *A priori* this is just a function on $\mathcal{H}$. Already we see, though, that $f|_{\Gamma g \Gamma'}$ only depends on the double coset $\Gamma g \Gamma'$ and not on the representatives g_i. Let $\gamma' \in \Gamma'$, then according to our decomposition of $\Gamma g \Gamma'$, for all i we have $g_i \gamma' = \gamma_i g_j$, for some $\gamma_i \in \Gamma$ and some index j. The map $i \mapsto j$ is a permutation of the indices. We find

$$(f|_{\Gamma g \Gamma'})\,|_{\gamma'} = \sum_i f|_{g_i \gamma'} = \sum_j f|_{\gamma_i g_j} = \sum_j f|_{g_j} = f|_{\Gamma g \Gamma}$$

Therefore $f|_{\Gamma g \Gamma}$ belongs to $M_k(\Gamma')$. The map

$$\begin{array}{rcl} M_k(\Gamma) & \to & M_k(\Gamma') \\ f & \mapsto & f|_{\Gamma g \Gamma'} \end{array}$$

is a Hecke correspondence.

In fact you get such an operator for every *function* h on $\mathrm{GL}_2^+(\mathbf{Q})$ which is left-invariant by Γ' and right-invariant by Γ, and which has finite support on $\mathrm{GL}_2^+(\mathbf{Q})/\Gamma$. Then h operators on modular forms f via *convolution*:

$$h \star f = \sum_{g \in \mathrm{GL}_2^+(\mathbf{Q})/\Gamma} h(g) f|_{g^{-1}}$$

Note that $f|_{g^{-1}}$ only depends on the coset of g in $\mathrm{GL}_2^+(\mathbf{Q})/\Gamma$, so that this expression makes sense, and that for $\gamma \in \Gamma'$, we have

$$\begin{aligned} (h \star f)|_\gamma &= \sum_{g \in \mathrm{GL}_2^+(\mathbf{Q})/\Gamma} h(g) f|_{g^{-1}\gamma} \\ &= \sum_{g \in \mathrm{GL}_2^+(\mathbf{Q})/\Gamma} h(\gamma g) f|_{g^{-1}} \\ &= h \star f \end{aligned}$$

because h is left Γ'-invariant.

The operator $f \mapsto f|_{\Gamma g \Gamma'}$ is the special case of when h is the characteristic function of $\Gamma' g^{-1} \Gamma$.

Suppose Γ'' is a third congruence subgroup. If h_1 is a function on $\Gamma' \backslash \mathrm{GL}_2^+(\mathbf{Q})/\Gamma$ and h_2 is a function on $\Gamma'' \backslash \mathrm{GL}_2^+(\mathbf{Q})/\Gamma'$, then we can define the convolution product $h_2 \star h_1$ by

$$(h_2 \star h_1)(x) = \sum_{g \in \mathrm{GL}_2^+(\mathbf{Q})/\Gamma'} h_2(g) h_1(g^{-1}x),$$

provided the sum is finite. Then

$$(h_2 \star h_1) \star f = h_2 \star (h_1 \star f).$$

If there is only one congruence subgroup Γ involved, then the $\mathbf{C}$-vector space of complex-valued functions $\mathcal{H}(\mathrm{GL}_2^+(\mathbf{Q}), \Gamma)$ which are left and right Γ-invariant and which are finitely supported on $\mathrm{GL}_2^+(\mathbf{Q})/\Gamma$. Then $\mathcal{H}(\mathrm{GL}_2^+(\mathbf{Q}))$ is an *algebra* under $\star$, called the Hecke algebra; the identity is the characteristic function of Γ itself. The convolution product $h \star f$ gives a representation of $\mathcal{H}(\mathrm{GL}_2^+(\mathbf{Q}), \Gamma)$ on the complex vector space $M_k(\Gamma)$. The Hecke algebra also preserves $S_k(\Gamma)$.

We examine the case of $\Gamma = \mathrm{SL}_2(\mathbf{Z})$. For a prime p, we write $\tilde{T}_p$ for the Hecke operator corresponding to the double coset $\Gamma \begin{pmatrix} 1 & \\ & p \end{pmatrix} \Gamma$. That is, for $f \in M_k(\Gamma)$ we write $\tilde{T}_p f$ for $f|_{\Gamma \mathrm{diag}(1,p) \Gamma}$.

We decompose the double coset space as

$$\Gamma \begin{pmatrix} 1 & \\ & p \end{pmatrix} \Gamma = \Gamma \begin{pmatrix} p & \\ & 1 \end{pmatrix} \cup \bigcup_{i=0}^{p} \Gamma \begin{pmatrix} 1 & i \\ & p \end{pmatrix}$$

Why does it work this way? An $g \in \Gamma \begin{pmatrix} 1 & \\ & p \end{pmatrix} \Gamma$ translates the standard basis of $\mathbf{Z}^2$ onto the basis of a sublattice $L \subset \mathbf{Z}^2$ of index p. Then L depends only on the image of g in $\Gamma \backslash \mathrm{GL}_2^+(\mathbf{Q})$. Conversely, L determines g. Thus the above double coset decomposition is equivalent to classifying the sublattices $L \subset \mathbf{Z}^2$ of index p. There are $p+1$ of these: the one generated by $(p, 0)$ and $(0, 1)$, and the ones generated by $(1, 0)$ and (a, p) for $a = 0, \ldots, p-1$.

For $f \in M_k(\Gamma)$ this works out to

$$\tilde{T}_p f(z) = p^{k/2} f(pz) + p^{-k/2} \sum_{a=0}^{p-1} f\left(\frac{z+a}{p}\right)$$

If f is a normalized cusp form with Fourier expansion

$$f(z) = \sum_{n \geq 1} a_n q^n, \; a_1 = 1$$

then

$$\tilde{T}_p f(z) = p^{k/2} \sum_{n \geq 1} a_n q^{pn} + p^{-k/2} \sum_{n \geq 1} a_n \sum_{b=0}^{p-1} \zeta_p^{bn} q^{n/p}$$

The sum over b is zero unless n is divisible by p, in which case it is p:

$$\tilde{T}_p f(z) = p^{k/2} \sum_{n \geq 1} a_n q^{pn} + p^{1-k/2} \sum_{n \geq 1} a_{pn} q^n$$

It is convenient to rescale the Hecke operator so that there aren't any denominators:

$$T_p = p^{\frac{k}{2}-1} \tilde{T}_p.$$

Then

$$T_p f(z) = p^{k-1} \sum_{n \geq 1} a_n q^{pn} + \sum_{n \geq 1} a_{pn} q^n \tag{3.1.1}$$

The operators T_n for n composite can be defined the same way, although the double coset decomposition will be more complicated. One has the rules

$$\begin{aligned} T_m T_n &= T_{mn}, \; \gcd(m, n) = 1 \\ T_p T_{p^n} &= T_{p^{n+1}} + p^{k-1} T_{p^{n-1}} \end{aligned}$$

In particular all the T_m commute with one another. One also checks that

$$(T_n f, g)_{k,\Gamma} = (f, T_n g)_{k,\Gamma}$$

so that the T_n are a collection of commuting self-adjoint operators on a finite-dimensional Hilbert space. Therefore they can be simultaneously diagonalized.

3.2 The L-function of a normalized eigenform

Let $f = \sum_{a_n} q^n$ be a normalized *eigenform* with $T_p f = \lambda_p f$. It follows immediately from our explicit formula for the action of T_p on q-expansions that $a_p = \lambda_p$. (In fact the eigenvalue of T_n is a_n for all n.) Thus f is completely determined by its sequence of Hecke eigenvalues.

We also have the estimate $a_n = O(n^{k/2})$ (see Gelbart Cor. 1.6). This implies that $f(z)$ approaches zero very rapidly as $\Im z \to \infty$; indeed $f(z) = O(e^{-2\pi y}$.

Since $T_m T_n = T_{mn}$ for relatively prime m, n, we have $a_m a_n = a_{mn}$. Thus the L-function of f admits an Euler factorization:

$$L(f,s) = \prod_p L_p(f,s)$$

with

$$L_p(f,s) = \sum_{n \geq 0} \frac{a_{p^n}}{p^{ns}}.$$

The recursion relating the T_{p^n} implies that

$$a_p a_{p^n} = a_{p^{n+1}} + p^{k-1} a_{p^{n-1}},$$

and with this we can determine $L_p(f,s)$:

$$\begin{aligned} \left(1 - \frac{a_p}{p^s} + \frac{p^{k-1}}{p^{2s}}\right) L_p(f,s) &= \sum_{n \geq 0} \frac{1}{p^{ns}} \left(a_{p^n} - a_p a_{p^{n-1}} + p^{k-1} a_{p^{n-2}}\right) \\ &= 1 \end{aligned}$$

(We set $a_r = 0$ for non-integral r. The expression in parentheses is 0 unless $n = 0$.) Thus

$$L_p(f,s) = \left(1 - \frac{a_p}{p^s} + \frac{p^{k-1}}{p^{2s}}\right)^{-1}$$

3.3 The Ramanujan-Petersson Conjecture

The cusp form of lowest weight for $\mathrm{SL}_2(\mathbf{Z})$ is

$$\Delta(z) = q \prod_{n \geq 1} (1 - q^n)^{24} = \sum_{n \geq 1} \tau(n) q^n.$$

This belongs to $S_{12}(\mathrm{SL}_2(\mathbf{Z}))$, which is one-dimensional. Thus Δ is automatically an eigenform. It follows that

$$L(\Delta, s) = \prod_p \left(1 - \frac{\tau(p)}{p^s} + \frac{p^{11}}{p^{2s}}\right)^{-1},$$

a fact which was conjectured by Ramanujan in 1916 and proved by Mordell in 1920. Ramanujan also conjectured that

$$|\tau(p)| \leq 2p^{11/2}.$$

Petersson's generalization of this is as follows: if $f = \sum_{n \geq 1} a_n q^n \in S_k(\Gamma_0(N))$ is a normalized eigenform, and p is a prime not dividing N, then

$$|a_p| \leq 2p^{(k-1)/2}$$

This inequality has a representation-theoretic significance which will become clear only later. For now we observe that if

$$L_p(f, s) = (1 - \alpha_p p^{-s})^{-1}(1 - \beta_p p^{-s})^{-1},$$

so that

$$\alpha_p, \beta_p = \frac{a_p \pm \sqrt{a_p^2 - 4p^{k-1}}}{2},$$

then the Ramanujan-Petersson conjecture holds if and only if the expression under the radical is non-positive, which is true if and only if

$$|\alpha_p| = |\beta_p| = p^{\frac{k-1}{2}}$$

Another reformulation is that the poles of $L_p(f, s)$ all have real part equal to $(k-1)/2$, which is a kind of local Riemann hypothesis for f. It isn't a superficial analogy: Deligne's 1974 proof of the Weil conjectures, which include the Riemann hypothesis for varieties over finite fields, has the Ramanujan-Petersson conjecture for cusp forms as a (difficult) corollary.

Finally, we mention that the RP conjecture says that the eigenvalues of the nonnormalized Hecke operator $\tilde{T}_p$ are bounded by $2\sqrt{p}$, a formulation which has the virtue of not depending on the weight k.

3.4 The L-function

Let f be a normalized eigenform in $S_k(\mathrm{SL}_2(\mathbf{Z}))$ (so that k is even). Since $a_n = O(n^{k/2})$, the Dirichlet series $L(f,s) = \sum_{n\geq 1} a_n n^{-s}$ only converges *a priori* for $\Re s > k/2+1$. Put

$$\Lambda(f,s) = (2\pi)^{-s}\Gamma(s)L(f,s) = \int_0^\infty f(iy)y^s \frac{dy}{y}.$$

Noting that $f(-1/y) = y^k f(y)$, we have

$$\begin{aligned}\Lambda(f,s) &= \int_1^\infty f(iy)y^s\frac{dy}{y} + \int_0^1 f(iy)y^s\frac{dy}{y} \\ &= \int_1^\infty f(iy)y^s\frac{dy}{y} + \int_1^\infty f(i/y)y^{-s}\frac{dy}{y} \\ &= \int_1^\infty f(iy)y^s\frac{dy}{y} + \int_1^\infty i^k f(iy)y^{k-s}\frac{dy}{y};\end{aligned}$$

this expression shows that $\Lambda(f,s)$ extends to an *entire* function of s satisfying

$$\Lambda(f,s) = i^k\Lambda(f,k-s).$$

The analytic number theorists prefer to normalize the L-function like this:

$$L_{\mathrm{an}}(f,s) = L\left(f, s+\frac{k-1}{2}\right)$$

This way, the functional equation relates $L_{\mathrm{an}}(f,s)$ to $L_{\mathrm{an}}(f,1-s)$.

3.5 Hecke's converse theorem

Suppose $\sum_{n\geq 1} a_n n^{-s}$ is a Dirichlet series which converges in some right half-plane to a function $L(s)$. Suppose that $L(s)$ behaves like the L-function attached to a modular form of weight k. That is, $L(s)$ is entire and $\Lambda(s) = (2\pi)^{-s}\Gamma(s)L(s)$ satisfies $\Lambda(s) = i^k\Lambda(k-s)$. Hecke showed that there exists a cusp form $f \in S_k(\mathrm{SL}_2(\mathbf{Z}))$ for which $L(s) = L(f,s)$. The proof is simple: the function $f(z) = \sum_{n\geq 1} a_n q^n$ is easily shown to satisfy $f(z+1) = f(z)$ (because it is a function of $q = e^{2\pi i z}$), and the functional equation of $L(s)$ shows that $f(-1/z) = z^k f(z)$. One then applies the fact that $\begin{pmatrix} 1 & 1 \\ & 1 \end{pmatrix}$ and $\begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$ generate $\mathrm{SL}_2(\mathbf{Z})$.

4 Hecke's theory II: Modifications to the theory in higher level

4.1 $\Gamma_0(N)$ and $\Gamma_1(N)$

The space of modular forms $S_k(\Gamma_0(N))$ arises very often, especially in the context of elliptic curves and modular abelian varieties. Sometimes it is just written $S_k(N)$. It won't be completely clear why the groups $\Gamma_0(N)$ and $\Gamma_1(N)$ deserve such status above other congruence subgroups, at least until we develop the representation-theoretic point of view. In the meantime it is important to know the basic modifications to Hecke's theory required in higher level, especially the phenomenon of *newforms*.

Note that $\Gamma_1(N) \subset \Gamma_0(N)$ is a normal subgroup, with quotient isomorphic to $(\mathbf{Z}/N\mathbf{Z})^\times$. Therefore the space $S_k(\Gamma_1(N))$ admits an action of $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbf{Z}/N\mathbf{Z})^\times$, via $f \mapsto f|_\gamma$. This action allows us to decompose $S_k(\Gamma_1(N))$ into eigenspaces $S_k(N,\chi)$ indexed by the Dirichlet characters modulo N:

$$S_k(\Gamma_1(N)) = \bigoplus_\chi S_k(N,\chi)$$

There are Hecke operators T_n acting on $S_k(N,\chi)$ for all n, but the formula for the action of (say) T_p on q-expansions is different from the one appearing in Eq. (3.1.1).

The Hecke operators also act differently with respect to the Petersson inner product on $S_k(N,\chi)$. When p does not divide N, the rule is

$$(T_p f, g) = \chi(p)(f, T_p g)$$

So there still exists a basis of $S_k(N,\chi)$ consisting of eigenforms for the T_p, $p \nmid N$. However, the situation really is different for $p|N$. For such p, T_p doesn't behave well at all with respect to (f,g), and it is no longer true that T_p can be diagonalized on $S_k(N,\chi)$. As a result, it is possible to have two distinct normalized forms with the same Hecke eigenvalues for *all* T_n. This isn't possible when $N = 1$!

4.2 Old and newforms

The situation can be remedied by dividing the new forms from the old. First consider the case of trivial character. For each proper divisor $m|N$, and each divisor d of N/m, we have a map

$$\begin{aligned} S_k(m) &\to S_k(N) && (4.2.1)\\ f(z) &\mapsto f(dz) && (4.2.2) \end{aligned}$$

The images of all such maps generate a subspace $S_k(N)^- \subset S_k(N)$, the *old subspace*. Eigenforms in $S_k(N)^-$ have levels which are strictly smaller than N. Let $S_k(N)^+$ be the subspace of $S_k(N)^-$ which is *complementary* to $S_k(N)^-$ under the Petersson inner product. This is the *new subspace*.

It is immediate that $S_k(N)^+$ is preserved by T_p, all $p \nmid N$, and that one can find a basis for $S_k(N)$ consisting of eigenforms f for these T_p. Call such an f a *newform* of level N. It is a nontrivial task to determine how the T_p for $p|N$ act on newforms. It turns out that for $p|N$, $T_p f = \lambda_p f$, where

$$\lambda_p = \begin{cases} \pm p^{k/2-1}, & p||N \\ 0, & p^2|N \end{cases}$$

In particular, newforms are eigenforms for *all* T_p.

Something very similar can be done to define newforms in $S_k(N, \chi)$, but here one observes that you only get old forms for those levels m for which χ descends to a Dirichlet character modulo m. (As an extreme case, if χ is a primitive Dirichlet character modulo N, then all eigenforms in $S_k(N, \chi)$ are automatically new.)

4.3 The L-functions of newforms: the Euler factors

Let $f = \sum_{n \geq 1} a_n q^n$ be a newform in $S_k(N, \chi)$. For $p \nmid N$, the p-Euler factor is

$$L_p(f, s) = \left(1 - \frac{a_p}{p^s} + \frac{p^{k-1}\chi(p)}{p^{2s}}\right)^{-1}.$$

If p divides N once, then

$$L_p(f, s) = \left(1 - \frac{a_p}{p^s}\right)^{-1}.$$

And if p^2 divides N, then

$$L_p(f, s) = 1.$$

4.4 The L-functions of newforms: the functional equation

Start with $\Gamma_0(N)$. We can't exactly replicate the derivation of the functional equation in the case of level 1, because $\begin{pmatrix} & 1 \\ -1 & \end{pmatrix}$ does not normalize $\Gamma_0(N)$ for $N > 1$. But $\begin{pmatrix} & -1 \\ N & \end{pmatrix}$ does normalize $\Gamma_0(N)$, so we can define an operator W_N on $S_k(N)$ by $W_N(f) = f|_{\begin{pmatrix} & -1 \\ N & \end{pmatrix}}$. Then $W_N^2 = 1$; W_N is called the *Atkin-Lehner* involution. It can be shown that W_N commutes with the Hecke operators T_p for $p \nmid N$, so that W_N takes a normalized newform f onto another newform with the same prime-to-N Hecke eigenvalues. But this must be a multiple of f itself: $W_N f = \varepsilon f$, for $\varepsilon = \pm 1$. One arrives at a functional equation of the form

$$\Lambda(f, s) = \varepsilon i^k \Lambda(f, k - s)$$

Things are a bit more complicated for forms with nontrivial character. If $f \in S_k(N, \chi)$ is a normalized newform with prime-to-N eigenvalues a_p, then $W_N f$ has prime-to-N eigenvalues $\chi(p)^{-1} a_p$; there exists a normalized newform $g \in S_k(N, \chi^{-1})$ and a constant w such that $W_N f = wg$. Then

$$\Lambda(f, s) = w i^k \Lambda(g, k - s).$$

This constant w, which Atkin and Li call a *pseudo-eigenvalue*, is shown to factor as a product of local constants, just as in Tate's theory. These constants really are quite subtle, especially in the cases where p^2 divides N. We won't be able to give a meaning to the local constants until we develop the adelic formulation of the theory of modular forms.