24 Modular forms and L-functions

As we will prove in the next lecture, Fermat's Last Theorem is a corollary of the following theorem for elliptic curves over $\mathbb{Q}$ [19, 20].

Theorem 24.1 (Taylor-Wiles)**.** *Every semistable elliptic curve $E/\mathbb{Q}$ is modular.*

In fact, as a result of subsequent work [4], we now have the following stronger result.

Theorem 24.2 (Breuil-Conrad-Diamond-Taylor)**.** *Every elliptic curve $E/\mathbb{Q}$ is modular.*

In this lecture we will explain what it means for an elliptic curve over $\mathbb{Q}$ to be modular (we will also define the term semistable).

This requires us to delve briefly into the theory of modular forms. Our goal in doing so is simply to understand the definitions and the terminology; we will omit all but the most straightforward proofs.

24.1 Modular forms

Definition 24.3. A holomorphic function $f\colon \mathcal{H} \to \mathbb{C}$ is a *weak modular form* of *weight* k for a congruence subgroup Γ if

$$f(\gamma\tau) = (c\tau + d)^k f(\tau)$$

for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$.

Example 24.4. The j-function $j(\tau)$ is a weak modular form of weight 0 for $\mathrm{SL}_2(\mathbb{Z})$, and $j(N\tau)$ is a weak modular form of weight 0 for $\Gamma_0(N)$. For an example of a weak modular form of positive weight, recall the Eisenstein series

$$G_k(\tau) := G_k([1,\tau]) := \sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}} \frac{1}{(m+n\tau)^k},$$

which, for $k \geq 3$, is a weak modular form of weight k for $\mathrm{SL}_2(\mathbb{Z})$. To see this, recall that $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, and note that

$$G_k(S\tau) = G_k(-1/\tau) = \sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}} \frac{1}{(m-\frac{n}{\tau})^k} = \sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}} \frac{\tau^k}{(m\tau-n)^k} = \tau^k G_k(\tau),$$

$$G_k(T\tau) = G_k(\tau+1) = G_k(\tau) = 1^k G_k(\tau).$$

If Γ contains $-I$, then any weak modular form f for Γ must satisfy $f(\tau) = (-1)^k f(\tau)$, since $-I$ acts trivially and $c\tau+d = -1$; this implies that when $-I \in \Gamma$ the only weak modular form of odd weight is the zero function. We are specifically interested in the congruence subgroup $\Gamma_0(N)$, which contains $-I$, so we will restrict our attention to modular forms of even weight, but we should note that for other congruence subgroups such as $\Gamma_1(N)$ that do not contain $-I$ (for $N > 2$) there are interesting modular forms of odd weight.

As we saw with modular functions (see Lecture 19), if Γ is a congruence subgroup of level N, meaning that it contains $\Gamma(N)$, then Γ contains the matrix $T^N = \left(\begin{smallmatrix} 1 & N \\ 0 & 1 \end{smallmatrix}\right)$, and every

weak modular form $f(\tau)$ for Γ must satisfy $f(\tau+N)=f(\tau)$ for $\tau\in\mathcal{H}$, since for T^N we have $(c\tau+d)^k=1^k=1$. It follows that $f(\tau)$ has a q-expansion of the form

$$f(\tau)=f^*(q^{1/N})=\sum_{n=-\infty}^{\infty}a_nq^{n/N}\qquad(q:=e^{2\pi i\tau}).$$

We say that f is *holomorphic at* ∞ if f^* is holomorphic at 0, equivalently, $a_n=0$ for $n<0$. We say that f is *holomorphic at the cusps* if $f(\gamma\tau)$ is holomorphic at ∞ for all $\gamma\in\mathrm{SL}_2(\mathbb{Z})$. As with modular functions, we only need to check this condition at a (finite) set of cusp representatives for Γ (if f is holomorphic at a particular cusp in $\mathbb{P}^1(\mathbb{Q})$ then it is necessarily holomorphic at every Γ-equivalent cusp). We should note that a weak modular form of positive weight is not Γ-invariant, so even when it is holomorphic on a cusp orbit, it may take on different values at cusps in the same orbit (but if it vanishes at a particular cusp then it vanishes at every Γ-equivalent cusp; this is relevant to the *cusp forms* defined below).

Definition 24.5. A *modular form* f is a weak modular form that is holomorphic at the cusps. Equivalently, f is a weak modular form that extends to a holomorphic function on the extended upper half plane $\mathcal{H}^*=\mathcal{H}\cup\mathbb{P}^1(\mathbb{Q})$.

If Γ is a congruence subgroup that contains the matrix $\left(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}\right)$ then every modular form for Γ has a q-series expansion at ∞ (or any cusp) of the form

$$f(\tau)=f^*(q)=\sum_{n\ge0}a_nq^n$$

that contains only integer powers of q, regardless of the level N. This includes the congruence subgroups $\Gamma_0(N)$ and $\Gamma_1(N)$ of interest to us. The coefficients a_n in the q-series for f are also referred to as the *Fourier coefficients* of f.

The only modular forms of weight 0 are constant functions. This is the main motivation for introducing the notion of weight, it allows us to generalize modular functions in an interesting way, by strengthening their analytic properties (holomorphic on $\mathcal{H}^*$, not just meromorphic) at the expense of weakening their congruence properties (modular forms of positive weight are not Γ-invariant due to the factor $(c\tau+d)^k$).

The j-function is not a modular form, since it has a pole at ∞, but the Eisenstein functions $G_k(\tau)$ are nonzero modular forms of weight k for $\mathrm{SL}_2(\mathbb{Z})$ for all even $k\ge4$. For $\Gamma=\mathrm{SL}_2(\mathbb{Z})$ there is only one cusp to check and it suffices to note that

$$\lim_{\mathrm{im}\,\tau\to\infty}G_k(\tau)=\lim_{\mathrm{im}(\tau)\to\infty}\sum_{\substack{m,n\in\mathbb{Z}\\(m,n)\neq(0,0)}}\frac{1}{(m+n\tau)^k}=2\sum_{n=1}^{\infty}\frac{1}{n^k}=2\zeta(k)<\infty,$$

(recall that the series converges absolutely, which justifies rearranging its terms).

Definition 24.6. A modular form is a *cusp form* if it vanishes at all the cusps. Equivalently, its q-expansion at every cusp has constant coefficient $a_0=0$.

Example 24.7. For even $k\ge4$ the Eisenstein series $G_k(\tau)$ is not a cusp form, but the discriminant function

$$\Delta(\tau)=g_2(\tau)^3-27g_3(\tau)^2,$$

with $g_2(\tau)=60G_4(\tau)$ and $g_3(\tau)=140G_6(\tau)$, is a cusp form of weight 12 for $\mathrm{SL}_2(\mathbb{Z})$; to see that it vanishes at ∞, note that $j(\tau)=g_2(\tau)^3/\Delta(\tau)$ has a pole at ∞ and $g_2(\tau)$ does not, so $\Delta(\tau)$ must vanish (see the proof of Theorem 15.11).

The set of modular forms of weight k for Γ is closed under addition and multiplication by constants $\lambda \in \mathbb{C}$ and thus forms a $\mathbb{C}$-vector space $M_k(\Gamma)$ that contains the cusp forms $S_k(\Gamma)$ as a subspace. We also note that if $f_1 \in M_{k_1}(\Gamma)$ and $f_2 \in M_{k_2}(\Gamma)$ then $f_1 f_2 \in M_{k_1+k_2}(\Gamma)$, but we will not use this fact.

Remarkably, the dimensions of the vector spaces $M_k(\Gamma)$ and $S_k(\Gamma)$ are finite, and can be explicitly computed in terms of invariants of the corresponding modular curve $X(\Gamma) = \mathcal{H}^*/\Gamma$.

As in Problem Set 10, let $\nu_2(\Gamma)$ count the number of Γ-inequivalent $\mathrm{SL}_2(\mathbb{Z})$-translates of i fixed by some $\gamma \in \Gamma$ other than $\pm I$ (elliptic points of period 2), and similarly define $\nu_3(\Gamma)$ in terms of $\rho = e^{2\pi i/3}$ (elliptic points of period 3). Let ν_∞ denote the number of cusp orbits, and let $g(\Gamma)$ be the genus of $X(\Gamma)$.

Theorem 24.8. *Let Γ be a congruence subgroup. For $k = 0$ we have $\dim M_k(\Gamma) = 1$ and $\dim S_k(\Gamma) = 0$. For any even integer $k > 0$ we have*

$$\dim M_k(\Gamma) = (k-1)(g(\Gamma)-1) + \left\lfloor \frac{k}{4} \right\rfloor \nu_2 + \left\lfloor \frac{k}{3} \right\rfloor \nu_3 + \frac{k}{2}\nu_\infty,$$

and if $k > 2$ we also have

$$\dim S_k(\Gamma) = (k-1)(g(\Gamma)-1) + \left\lfloor \frac{k}{4} \right\rfloor \nu_2 + \left\lfloor \frac{k}{3} \right\rfloor \nu_3 + \left(\frac{k}{2} - 1\right)\nu_\infty.$$

For $k = 2$ we have $\dim S_k(\Gamma) = g(\Gamma)$.

Proof. See [6, Thm. 3.5.1] □

We are specifically interested in the vector space $S_2(\Gamma_0(N))$ of dimension $g(\Gamma_0(N))$.

Remark 24.9. Those who know a bit of algebraic geometry may suspect that there is a relationship between the space of cusp forms $S_2(\Gamma_0(N))$ and the space of regular differentials for the modular curve $X_0(N)$, since their dimensions coincide; this is indeed the case.

24.2 Hecke operators

In order to understand the relationship between modular forms and elliptic curves we want to construct a canonical basis for $S_2(\Gamma_0(N))$. To help with this, we now introduce the *Hecke operators* T_n on $M_k(\Gamma_0(N))$; these are linear operators that fix the subspace $S_k(\Gamma_0(N))$.[1]

In order to motivate the definition of the Hecke operators on modular forms, we first define them in terms of lattices, following the presentation in [13, VII.5.1]. As in previous lectures, a lattice (in $\mathbb{C}$) is an additive subgroup of $\mathbb{C}$ that is a free $\mathbb{Z}$-module of rank 2 containing an $\mathbb{R}$-basis for $\mathbb{C}$.

For each positive integer n, the Hecke operator T_n sends each lattice $L = [\omega_1, \omega_2]$ to the formal sum of its index-n sublattices.

$$T_n L := \sum_{[L:L']=n} L'. \tag{1}$$

Here we are working in the free abelian group $\operatorname{Div}\mathcal{L}$ generated by the set $\mathcal{L}$ of all lattices; we extend T_n linearly to an endomorphism of $\operatorname{Div}\mathcal{L}$ (this means $T_n \sum L := \sum T_n L$). Another family of endomorphisms of $\operatorname{Div}\mathcal{L}$ are the homothety operators R_λ defined by

$$R_\lambda L := \lambda L, \tag{2}$$

[1] One can define Hecke operators more generally on $M_k(\Gamma_1(N))$, which contains $M_k(\Gamma_0(N))$, but the definition is more involved and not needed here.

for any $\lambda \in \mathbb{C}^\times$. This setup might seem overly abstract, but it allows one to easily prove some essential properties of the Hecke operators that are applicable in many settings. When defined in this generality the Hecke operators are also sometimes called *correspondences*.

Remark 24.10. Recall that if $E/\mathbb{C}$ is the elliptic curve isomorphic to the torus $\mathbb{C}/L$, the index-n sublattices of L correspond to n-isogenous elliptic curves. The fact that the Hecke operators average over sublattices is related to the fact that the relationship between modular forms and elliptic curves occurs at the level of isogeny classes.

Theorem 24.11. *The operators T_n and R_λ satisfy the following:*

(i) $T_n R_\lambda = R_\lambda T_n$ *and* $R_\lambda R_\mu = R_{\lambda\mu}$.

(ii) $T_{mn} = T_m T_n$ *for all* $m \perp n$.

(iii) $T_{p^{r+1}} = T_{p^r} T_p - p T_{p^{r-1}} R_p$ *for all primes p and integers $r \geq 1$.*

Proof. (i) is clear, as is (ii) if we note that for $m \perp n$ there is a bijection between index-mn sublattices L'' of L and pairs (L', L'') with $[L : L'] = n$ and $[L' : L''] = m$. For (iii), the first term on the RHS counts pairs (L', L'') with $[L : L'] = p$ and $[L' : L''] = p^r$, and the second term corrects for over counting; see [13, Prop. VII.10] for details. □

Corollary 24.12. *The subring of* $\mathrm{End}(\mathrm{Div}\,\mathcal{L})$ *generated by* $\{R_p, T_p : p \text{ prime}\}$ *is commutative and contains all the Hecke operators T_n.*

Proof. By recursively applying (iii) we can reduce any T_{p^r} to a polynomial in T_p and R_p, and any two such polynomials commute (since T_p and R_p commute, by (i)). Moreover, (i) and (ii) imply that for distinct primes p and q, polynomials in T_p, R_p commute with polynomials in T_q, R_q. Using (ii) and (iii) we can reduce any T_n to a product of polynomials in T_{p_i}, R_{p_i} for distinct primes p_i and the corollary follows. □

Any function $F\colon \mathcal{L} \to \mathbb{C}$ extends linearly to a function $F\colon \mathrm{Div}\,\mathcal{L} \to \mathbb{C}$ to which we may apply any operator $T \in \mathrm{End}(\mathrm{Div}\,\mathcal{L})$, yielding a new function $TF\colon \mathrm{Div}\,\mathcal{L} \to \mathbb{C}$ defined by $TF\colon D \mapsto F(T(D))$; restricting TF to $\mathcal{L} \subseteq \mathrm{Div}\,\mathcal{L}$ then gives a function $TF\colon \mathcal{L} \to \mathbb{C}$ that we regard as the transform of our original function F by T. This allows us to apply the Hecke operators T_n and homothety operators R_λ to any function that maps lattices to complex numbers. We will work this out explicitly for the Hecke operators acting on modular forms for $\mathrm{SL}_2(\mathbb{Z})$ in the next section.

24.3 Hecke operators for modular forms of level one

We now define the action of the Hecke operators T_n on $M_k(\mathrm{SL}_2(\mathbb{Z})) = M_k(\Gamma_0(1))$. The case $M_k(\Gamma_0(N))$ is analogous, but the details are more involved, so let us assume $N = 1$ for the sake of presentation and address $N > 1$ in remarks.

Let $f\colon \mathcal{H} \to \mathbb{C}$ be a modular form of weight k. We can view $f(\tau)$ as a function on lattices $[1, \tau]$, which we extend to arbitrary lattices $L = [\omega_1, \omega_2]$ by defining

$$f([\omega_1, \omega_2]) := f(\omega_1[1, \omega_2/\omega_1]) := \omega_1^{-k} f([1, \omega_2/\omega_1]),$$

we assume ω_1 and ω_2 are ordered so that ω_2/ω_1 is in the upper half plane. Conversely, any function $F\colon \mathcal{L} \to \mathbb{C}$ on lattices induces a function $\tau \mapsto F([1, \tau])$ on the upper half plane. Viewing our modular form f as a function $\mathcal{L} \to \mathbb{C}$, we can transform this function by any

$T \in \mathrm{End}(\mathrm{Div}\,\mathcal{L})$ as described above, thereby obtaining a new function $\mathcal{L} \to \mathbb{C}$ that induces a function $Tf\colon \mathcal{H} \to \mathbb{C}$ on the upper half plane. In general the function Tf need not be a modular form, but for $f \in M_k(\Gamma_0(1))$ it is (we will verify this in the cases of interest to us).

Motivated by the discussion above, for $f \in M_k(\Gamma_0(1))$ we define

$$R_\lambda f(\tau) := f(\lambda[1,\tau]) = \lambda^{-k} f(\tau),$$

which clearly lies in $M_k(\Gamma_0(1))$, and if f is a cusp form then so is $R_\lambda f$.

We define $T_n f$ similarly, but introduce a scaling factor of n^{k-1} that simplifies the formulas that follow. An easy generalization of Lemma 20.2 shows that for each integer $n \geq 1$, the index n sublattices of $[1,\tau]$ are given by

$$\Big\{[d, a\tau + b] : ad = n,\ 0 \leq b < d\Big\};$$

see [13, Lem. VII.5.2], for example. If we rescale by d^{-1} to put them in the form $[1,\omega]$, we have $\omega = (a\tau + b)/d$. For $f \in M_k(\Gamma_0(1))$ we thus define $T_n f$ as

$$T_n f(\tau) := n^{k-1} \sum_{[[1,\tau]:L]=n} f(L) = n^{k-1} \sum_{ad=n,\, 0\leq b<d} d^{-k} f\left(\frac{a\tau+b}{d}\right),$$

which is also clearly an element of $M_k(\Gamma_0(1))$, and if f is a cusp form, so is $T_n f$. It is clear from the definition that T_n acts linearly, so it is a linear operator on the vector spaces $M_k(\Gamma_0(1))$ and $S_k(\Gamma_0(1))$. Theorem 24.11 then yields the following corollary.

Corollary 24.13. *The Hecke operators T_n for $M_k(\Gamma_0(1))$ satisfy $T_{mn} = T_m T_n$ for $m \perp n$ and $T_{p^{r+1}} = T_{p^r} T_p - p^{k-1} T_{p^{r-1}}$ for p prime.*

Proof. The first equality is clear; the second term on the RHS of the second equality arises from the fact that $pT_{p^{r-1}} R_p f = p^{k-1} T_{p^{r-1}} f$. □

The corollary implies that we may restrict our attention to the Hecke operators T_p for p prime. Let us compute the q-series expansion of $T_p f$, where $f(\tau) = \sum_{n=1}^\infty a_n q^n$ is a cusp form of weight k for $\Gamma_0(1)$. We have

$$\begin{aligned}
T_p f(\tau) &= p^{k-1} \sum_{\substack{ad=p\\0\leq b<d}} d^{-k} f\left(\frac{a\tau+b}{d}\right)\\
&= p^{k-1} f(p\tau) + p^{-1} \sum_{b=0}^{p-1} f\left(\frac{\tau+b}{p}\right)\\
&= p^{k-1} \sum_{n=1}^\infty a_n e^{2\pi i n p\tau} + p^{-1} \sum_{b=0}^{p-1} \sum_{n=1}^\infty a_n e^{2\pi i n(\tau+b)/p}\\
&= p^{k-1} \sum_{n=1}^\infty a_n q^{np} + p^{-1} \sum_{b=0}^{p-1} \sum_{n=1}^\infty a_n \zeta_p^{bn} q^{n/p}\\
&= p^{k-1} \sum_{n=1}^\infty a_{n/p} q^n + p^{-1} \sum_{n=1}^\infty a_n \left(\sum_{b=0}^{p-1} \zeta_p^{bn}\right) q^{n/p}\\
&= \sum_{n=1}^\infty \left(a_{np} + p^{k-1} a_{n/p}\right) q^n,
\end{aligned}$$

where $\zeta_p = e^{2\pi i/p}$ and $a_{n/p}$ is defined to be 0 when $p \nmid n$. This calculation yields the following theorem and corollary, in which we use $a_n(f)$ to denote the coefficient of q^n in the q-expansion of f.

Theorem 24.14. *For any $f \in S_k(\Gamma_0(1))$ and prime p we have*

$$a_n(T_p f) = \begin{cases} a_{np}(f) & \text{if } p \nmid n, \\ a_{np}(f) + p^{k-1} a_{n/p}(f) & \text{if } p \mid n. \end{cases}$$

Corollary 24.15. *For any modular form $f \in S_k(\Gamma_0(1))$ and integers $m \perp n$ we have $a_m(T_n f) = a_{mn}(f)$; in particular, $a_1(T_n f) = a_n(f)$.*

Proof. The corollary follows immediately from Theorem 24.14 for n prime. For composite n (and any $m \perp n$), we proceed by induction on n. If $n = cd$ with $c \perp d$ both greater than 1, then by Theorem 24.14 and the inductive hypothesis we have

$$a_m(T_n f) = a_m(T_c T_d f) = a_{mc}(T_d f) = a_{mcd} = a_{mn}.$$

For $n = p^{r+1}$, applying Theorem 24.14, Corollary 24.13, and the inductive hypothesis yields

$$\begin{aligned} a_m(T_{p^{r+1}} f) &= a_m(T_{p^r} T_p f) - p^{k-1} a_m(T_{p^{r-1}} f) \\ &= a_{mp^r}(T_p f) - p^{k-1} a_{mp^{r-1}}(f) \\ &= a_{mp^{r+1}}(f) + p^{k-1} a_{mp^{r-1}}(f) - p^{k-1} a_{mp^{r-1}}(f) \\ &= a_{mn}(f), \end{aligned}$$

as desired. □

Remark 24.16. All the results in this section hold for $f \in S_k(\Gamma_0(N))$ if we restrict to Hecke operators T_n with $n \perp N$, which is all that we require, and the key result $a_1(T_n f) = a_n(f)$ holds in general. For $p|N$ the definition of T_p (and T_n for $p|n$) needs to change and the formulas in Corollary 24.13 and Theorem 24.14 must be modified. The definition of the Hecke operators is more complicated (in particular, it depends on the level N), but some of the formulas are actually simpler (for example, for $p|N$ we have $T_{p^r} = T_p^r$).

24.4 Eigenforms for the Hecke operators

The Hecke operators T_n defined in the previous section form an infinite family of linear operators on the vector space $S_k(\Gamma_0(1))$. We are interested in the elements $f \in S_k(\Gamma_0(1))$ that are simultaneous eigenvectors for all the Hecke operators; this means that for every $n \geq 1$ we have $T_n f = \lambda_n f$ for some eigenvalue $\lambda_n \in \mathbb{C}$ of T_n. When such an f also satisfies $a_1(f) = 1$, we call it a (normalized) *eigenform*. It is not immediately obvious that such f exist, but we will prove that they do, and that they provide a canonical basis for $S_k(\Gamma_0(1))$.

Given an eigenform f, we can read off the corresponding Hecke eigenvalues λ_n from its q-expansion $f = \sum a_n q^n$: if $T_n f = \lambda_n f$ then we must have

$$\lambda_n = \lambda_n a_1 = a_1(T_n f) = a_n(f) = a_n,$$

by Corollary 24.15. Corollary 24.13 implies that the a_n then satisfy

$$\begin{aligned} a_{mn} &= a_m a_n && (m \perp n), \\ a_{p^r} &= a_p a_{p^{r-1}} - p^{k-1} a_{p^{r-2}} && (p \text{ prime}). \end{aligned} \tag{3}$$

In particular, the coefficients a_n are completely determined by the values a_p at primes p.

Remark 24.17. For $k = 2$ the recurrence for a_{p^r} should look familiar: it is the same recurrence satisfied by the Frobenius traces $a_{p^r} := p^r + 1 - \#E(\mathbb{F}_{p^r})$ of an elliptic curve $E/\mathbb{F}_p$, as shown in Problem Set 7.

Our goal in this section is to construct a basis of eigenforms for $S_k(\Gamma_0(1))$, and prove that it is unique. In order to do so, we need to introduce the *Petersson inner product*, which defines a Hermitian form on the $\mathbb{C}$-vector spaces $S_k(\Gamma)$ (for any congruence subgroup Γ). Recall that for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$, we have $\operatorname{im}\gamma\tau = \operatorname{im}\tau/|c\tau+d|^2$, thus for any $f, g \in S_k(\Gamma)$ we have

$$f(\gamma\tau)\overline{g(\gamma\tau)}(\operatorname{im}\gamma\tau)^k = (c\tau+d)^k f(\tau)(c\bar{\tau}+d)^k g(\tau)\left(\frac{\operatorname{im}\tau}{|c\tau+d|^2}\right)^k = f(\tau)\overline{g(\tau)}(\operatorname{im}\tau)^k.$$

The function $f(\tau)\overline{g(\tau)}(\operatorname{im}\tau)^k$ is thus Γ-invariant. If we parameterize the upper half-plane $\mathcal{H}$ with real parameters $x = \operatorname{re}\tau$ and $y = \operatorname{im}\tau$, so $\tau = x + iy$, it is straight-forward to check that the measure

$$\mu(U) = \iint_U \frac{dxdy}{y^2}$$

is $\mathrm{SL}_2(\mathbb{Z})$-invariant (hence Γ-invariant), that is, $\mu(\gamma U) = \mu(U)$ for all measurable sets $U \subseteq \mathcal{H}$. This motivates the following definition.

Definition 24.18. The *Petersson inner product* on $S_k(\Gamma)$ is defined by

$$\langle f, g\rangle = \int_{\mathcal{F}} f(\tau)\overline{g(\tau)}y^{k-2}dxdy, \tag{4}$$

where the integral ranges over points $\tau = x + yi$ in a fundamental region $\mathcal{F} \subseteq \mathcal{H}$ for Γ. It is easy to check that $\langle f, g\rangle$ is a positive definite Hermitian form: it is sesquilinear in f and g, it satisfies $\langle f, g\rangle = \overline{\langle g, f\rangle}$, and $\langle f, f\rangle \geq 0$ with equality only when $f = 0$. It thus defines an inner product on the $\mathbb{C}$-vector space $S_k(\Gamma)$.

One can show that the Hecke operators for $S_k(\Gamma_0(1))$ are self-adjoint with respect to the Petersson inner product, that is, they satisfy $\langle f, T_n g\rangle = \langle T_n f, g\rangle$. The T_n are thus Hermitian (normal) operators, and we know from Corollary 24.13 that they all commute with each other. This makes it possible to apply the following form of the Spectral Theorem.

Lemma 24.19. *Let V be a finite-dimensional $\mathbb{C}$-vector space equipped with a positive definite Hermitian form, and let $\alpha_1, \alpha_2, \ldots$ be a sequence of commuting Hermitian operators. Then $V = \bigoplus_i V_i$, where each V_i is an eigenspace of every α_n.*

Proof. The matrix for α_1 is Hermitian, therefore diagonalizable,[2] so we can decompose V as a direct sum of eigenspaces for α_1, writing $V = \bigoplus_i V(\lambda_i)$, where the λ_i are the distinct eigenvalues of α_1. Because α_1 and α_2 commute, α_2 must fix each subspace $V(\lambda_i)$, since for each $v \in V(\lambda_i)$ we have $\alpha_1\alpha_2 v = \alpha_2\alpha_1 v = \alpha_2\lambda_i v = \lambda_i\alpha_2 v$, and therefore $\alpha_2 v$ is an eigenvector for α_1 with eigenvalue λ_i, so $\alpha_2 v \in V(\lambda_i)$. Thus we can decompose each $V(\lambda_i)$ as a direct sum of eigenspaces for α_2, and may continue in this fashion for all the α_n. □

By Lemma 24.19, we may decompose $S_k(\Gamma_0(1)) = \bigoplus_i V_i$ as a direct sum of eigenspaces for the Hecke operators T_n. Let $f(\tau) = \sum a_n q^n$ be a nonzero element of V_i. We then have $a_1(T_n f) = a_n$, by Corollary 24.13, and also $T_n f = \lambda_n f$, for some eigenvalue λ_n of T_n which

[2] This fact is also sometimes called the Spectral Theorem and proved in most linear algebra courses.

is determined by V_i, so $a_n = \lambda_n a_1$. This implies $a_1 \neq 0$, since otherwise $f = 0$, and if we normalize f so that $a_1 = 1$ (which we can do, since f is nonzero and V_i is a $\mathbb{C}$-vector space), we then have $a_n = \lambda_n$ for all $n \geq 1$, and f completely determined by the sequence of Hecke eigenvalues λ_n for V_i. It follows that every element of V_i is a multiple of f, so $\dim V_i = 1$ and the eigenforms in $S_k(\Gamma_0(1))$ form a basis.

Theorem 24.20. *The space of cusp forms $S_k(\Gamma_0(1))$ is a direct sum of one-dimensional eigenspaces for the Hecke operators T_n and has a unique basis of eigenforms $f(\tau) = \sum a_n q^n$, where each a_n is the eigenvalue of T_n on the one-dimensional subspace spanned by f.*

The analog of Theorem 24.20 fails for $S_k(\Gamma_0(N))$ for two reasons, both of which are readily addressed. First, as in Remark 24.16, we need to restrict our attention to the Hecke operators T_n with $n \perp N$ (when n and N have a common factor T_n is not necessarily a Hermitian operator with respect to the Petersson inner product). We can then proceed as above to decompose $S_k(\Gamma_0(N))$ into eigenspaces for the Hecke operators T_n with $n \perp N$. We then encounter the second issue, which is that these eigenspaces need not be one-dimensional. In order to obtain a decomposition into one-dimensional eigenspaces we must restrict our attention to a particular subspace of $S_k(\Gamma_0(N))$.

Note that for any $M|N$ the space $S_k(\Gamma_0(M))$ is a subspace of $S_k(\Gamma_0(N))$ (since $\Gamma_0(M)$-invariance implies $\Gamma_0(N)$-invariance for $M|N$). We say that a cusp form $f \in S_k(\Gamma_0(N))$ is *old* if it also lies in the subspace $S_k(\Gamma_0(M))$ for some M properly dividing N. The oldforms in $S_k(\Gamma_0(N))$ generate a subspace $S_k^{\text{old}}(\Gamma_0(N))$, and we define $S_k^{\text{new}}(\Gamma_0(N))$ as the orthogonal complement of $S_k^{\text{old}}(\Gamma_0(N))$ in $S_k(\Gamma^0(N))$ (with respect to the Petersson inner product), so that

$$S_k(\Gamma_0(N)) = S_k^{\text{old}}(\Gamma_0(N)) \oplus S_k^{\text{new}}(\Gamma_0(N)),$$

and we call the eigenforms in $S_k^{\text{new}}(\Gamma_0(N))$ *newforms* (normalized so $a_1 = 1$). One can show that the Hecke operators T_n with $n \perp N$ preserve both $S_k^{\text{old}}(\Gamma_0(N))$ and $S_k^{\text{new}}(\Gamma_0(N))$. If we then decompose $S_k^{\text{new}}(\Gamma_0(N))$ into eigenspaces with respect to these operators, the resulting eigenspaces are all one-dimensional, moreover, each is actually generated by an eigenform (a simultaneous eigenvector for *all* the T_n, not just those with $n \perp N$ that we used to obtain the decomposition); this is a famous result of Atkin and Lehner [3, Thm. 5]. Note that $S_k^{\text{new}}(\Gamma_0(1)) = S_k(\Gamma_0(1))$, and we thus have the following generalization of Theorem 24.20.

Theorem 24.21. *The space $S_k^{\text{new}}(\Gamma_0(N))$ is a direct sum of one-dimensional eigenspaces for the Hecke operators T_n and has a unique basis of newforms $f(\tau) = \sum a_n q^n$, where each a_n is the eigenvalue of T_n on the one-dimensional subspace spanned by f.*

24.5 The *L*-function of a modular form

Our interest in cusp forms is that each has an associated *L-function*, which is defined in terms of a particular *Dirichlet series*.

Definition 24.22. A *Dirichlet series* is a series of the form

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

where the a_n are complex numbers and s is a complex variable. Provided the a_n satisfy a polynomial growth bound of the form $|a_n| = O(n^\sigma)$ (as $n \to \infty$), the series $L(s)$ converges locally uniformly in the right half plane $\text{Re}(s) > 1 + \sigma$ and defines a holomorphic function

in this region (which may extend to a holomorphic or meromorphic function on a larger region).

Example 24.23. The most famous Dirichlet series is the *Riemann zeta function*

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}$$

which converges locally uniformly to a holomorphic function on $\mathrm{re}(s) > 1$. It has three properties worth noting:

- **analytic continuation**: $\zeta(s)$ extends to a meromorphic function on $\mathbb{C}$ (with a simple pole at $s = 1$ and no other poles);
- **functional equation**: the *completed zeta function*[3] $\hat{\zeta}(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)$ satisfies
$$\hat{\zeta}(s) = \hat{\zeta}(1-s);$$
- **Euler product**: we can write $\zeta(s)$ as a product over primes (for $\mathrm{re}(s) > 1$) via
$$\zeta(s) = \prod_p (1-p^{-s})^{-1} = \prod_p (1 + p^{-s} + p^{-2s} + \dots) = \sum_{n=1}^{\infty} n^{-s}.$$

Definition 24.24. The *L-function* (or *L-series*) of a cusp form $f(\tau) = \sum_{n=1}^{\infty} a_n q^n$ of weight k is the complex function defined by the Dirichlet series

$$L(f,s) := \sum_{n=1}^{\infty} a_n n^{-s},$$

which converges locally uniformly to a holomorphic function on $\mathrm{re}(s) > 1 + k/2$.

Theorem 24.25 (Hecke)**.** *Let $f \in S_k(\Gamma_0(N))$. The L-function $L(f,s)$ extends analytically to a holomorphic function on $\mathbb{C}$, and the normalized L-function*

$$\tilde{L}_f(s) = N^{s/2}(2\pi)^{-s}\Gamma(s)L(f,s)$$

satisfies the functional equation

$$\tilde{L}_f(s) = \pm\tilde{L}_f(k-s).$$

Remark 24.26. There are more explicit versions of this theorem that also determine the sign in the functional equation above.

For newforms we also get an Euler product.

Theorem 24.27. *Let $f \in S_k^{\mathrm{new}}(\Gamma_0(N))$. The L-function $L(f,s)$ has the Euler product*

$$L(f,s) = \sum_{n=1}^{\infty} a_n n^{-s} = \prod_p (1 - a_p p^{-s} + \chi(p)p^{k-1}p^{-2s})^{-1}, \tag{5}$$

where $\chi(p) = 0$ for $p|N$ and $\chi(p) = 1$ otherwise.

The function χ in Theorem 24.27 is the principal Dirichlet character of conductor N, a periodic function $\mathbb{Z} \to \mathbb{C}$ supported on $(\mathbb{Z}/N\mathbb{Z})^\times$ that defines a group homomorphism $(\mathbb{Z}/N\mathbb{Z})^\times \to \mathbb{C}$ (the adjective "principal" indicates that the homomorphism is trivial).

[3]Here $\Gamma(s) := \int_0^\infty e^{-t}t^{s-1}dt$ is Euler's gamma function.

24.6 The L-function of an elliptic curve

What does all this have to do with elliptic curves? Like eigenforms, elliptic curves over $\mathbb{Q}$ also have an L-function with an Euler product. In fact, with elliptic curves, we use the Euler product to define the L-function.

Definition 24.28. The *L-function of an elliptic curve* $E/\mathbb{Q}$ is given by the Euler product

$$L(E,s) = \prod_p L_p(p^{-s})^{-1} = \prod_p \left(1 - a_p p^{-s} + \chi(p) p p^{-2s}\right)^{-1}, \tag{6}$$

where $\chi(p)$ is 0 if E has bad reduction at p, and 1 otherwise.[4] For primes p where E has good reduction (all but finitely many), $a_p := p + 1 - \#E_p(\mathbb{F}_p)$ is the trace of Frobenius, where E_p denotes the reduction of E modulo p. Equivalently, $L_p(T)$ is the numerator of the zeta function

$$Z(E_p;T) = \exp\left(\sum_{n=1}^{\infty} \#E_p(\mathbb{F}_{p^n})\frac{T^n}{n}\right) = \frac{1 - a_p T + pT^2}{(1-T)(1-pT)},$$

that appeared in the special case of the Weil conjectures that you proved in Problem Set 7. For primes p where E has bad reduction, the polynomial $L_p(T)$ is defined by

$$L_p(T) = \begin{cases} 1 & \text{if } E \text{ has } \textit{additive} \text{ reduction at } p. \\ 1 - T & \text{if } E \text{ has } \textit{split multiplicative} \text{ reduction at } p. \\ 1 + T & \text{if } E \text{ has } \textit{non-split multiplicative} \text{ reduction at } p. \end{cases}$$

according to the type of bad reduction E has at p, as explained in the next section. This means that $a_p \in \{0, \pm 1\}$ at bad primes.

The L-function $L(E,s)$ converges to a holomorphic function on $\mathrm{re}(s) > 3/2$.

24.7 The reduction type of an elliptic curve

When computing $L(E,s)$, it is important to use a *minimal* Weierstrass equation for E, one that has good reduction at as many primes as possible. To see why this is necessary, note that if $y^2 = x^3 + Ax + B$ is a Weierstrass equation for E, then, up to isomorphism, so is $y^2 + u^4 Ax + u^6 B$, for any integer u, and this equation will have bad reduction at all primes $p|u$. Moreover, even though the equation $y^2 = x^3 + Ax + B$ always has bad reduction at 2, there may be an equation for E in general Weierstrass form that has good reduction at 2. For example, the elliptic curve defined by $y^2 = x^3 + 16$ is isomorphic to the elliptic curve defined by $y^2 + y = x^3$ (replace x by $4x$, divide by 64, and then replace y by $y + 1/2$), which does have good reduction at 2.

Definition 24.29. Let $E/\mathbb{Q}$ be an elliptic curve. A (global) *minimal model* for E is an integral Weierstrass equation

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6,$$

with $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$ that defines an elliptic curve isomorphic to E whose discriminant $\Delta_{\min}(E)$ divides the discriminant of every integral Weierstrass equation for E.

[4] As explained in §24.7, this assumes we are using a minimal Weierstrass equation for E.

It is not immediately obvious that minimal models necessarily exist, but for elliptic curves over $\mathbb{Q}$ this is so; see [16, Prop. VII.1.3].[5] One can construct a minimal model in Sage using `E.minimal_model()`; see [9] for an explicit algorithm.

We now address the three types of bad reduction. To simplify the presentation, we will ignore the prime 2, but the three cases described below also occur at 2. For any odd prime p of bad reduction we can represent the singular curve $E_p/\mathbb{F}_p$ by an equation of the form $y^2 = f(x)$, for some cubic $f \in \mathbb{F}_p[x]$ that has a repeated root r. The repeated root r is necessarily rational, and by replacing x with $x-r$ we can assume $r = 0$, so $y^2 = x^3 + ax^2$ for some $a \in \mathbb{F}_p$. The projective curve $y^2z = x^3 + ax^2z$ has exactly one singular point $(0 : 0 : 1)$ and is smooth elsewhere (including the point $(0 : 1 : 0)$ at infinity).

If we exclude the singular point $(0 : 0 : 1)$, the standard formulas for the group law on $E_p(\mathbb{F}_p)$ still make sense, and the set

$$E_p^{\mathrm{ns}}(\mathbb{F}_p) := E_p(\mathbb{F}_p) - \{(0 : 0 : 1)\}$$

of non-singular points of $E_p(\mathbb{F}_p)$ is closed under the group operation.[6] Thus $E_p^{\mathrm{ns}}(\mathbb{F}_p)$ is a finite abelian group. We now define

$$a_p := p - \#E_p^{\mathrm{ns}}(\mathbb{F}_p).$$

This is analogous to the good reduction case in which $a_p = p + 1 - \#E_p(\mathbb{F}_p)$; we have removed the (necessarily rational) singular point, so we reduce a_p by one.

There are two cases to consider, depending on whether $f(x)$ has a double or triple root at 0; these two cases give rise to three possibilities for the group $E_p^{\mathrm{ns}}(\mathbb{F}_p)$.

- **Case 1: triple root** ($y^2 = x^3$)

  We have the projective curve $zy^2 = x^3$. After removing the singular point $(0 : 0 : 1)$, every other projective point has non-zero y coordinate, so we can fix $y = 1$, and work with the affine curve $z = x^3$. There are p solutions to this equation (including $x = 0$ and $z = 0$, which corresponds to the projective point $(0 : 1 : 0)$ at infinity). It follows that $E_p^{\mathrm{ns}}(\mathbb{F}_p)$ is a cyclic group of order p, which is necessarily isomorphic to the additive group of $\mathbb{F}_p$; see [18, §2.10] for an explicit isomorphism. In this case we have $a_p = 0$ and say that E has *additive reduction* at p.

- **Case 2: double root** ($y^2 = x^3 + ax^2$, $a \neq 0$).

  We have the projective curve $zy^2 = x^3 + ax^2z$, and the point $(0 : 1 : 0)$ at infinity is the only non-singular point on the curve whose x or z coordinate is zero. Excluding the point at infinity for the moment, let us divide both sides by x^2, introduce the variable $t = y/x$, and fix $z = 1$. This yields the affine curve $t^2 = x + a$, and the number of

[5] For an elliptic curve E over a number field K one defines $\Delta_{\min}(E)$ as the $\mathcal{O}_K$-ideal generated by the discriminants of all integral models for E (with $a_1, a_2, a_3, a_4, a_6 \in \mathcal{O}_K$); if the class number of $\mathcal{O}_K$ is greater than one this ideal need not be a principal ideal, in which case E cannot have a minimal model over K.

[6] To see this geometrically, note that any line in $\mathbb{P}^2$ intersecting a plane cubic in two non-singular points cannot also intersect it in a singular point; when we count intersections with multiplicity the total must be three, by Bezout's theorem, but singular points contribute multiplicity greater than one.

points with $x \neq 0$ is

$$\begin{aligned}\sum_{x\neq 0}\left(1+\left(\frac{x+a}{p}\right)\right) &= \sum_{x}\left(1+\left(\frac{x+a}{p}\right)\right)-\left(1+\left(\frac{a}{p}\right)\right)\\ &= \sum_{x}\left(1+\left(\frac{x}{p}\right)\right)-1-\left(\frac{a}{p}\right)\\ &= p-1-\left(\frac{a}{p}\right)\end{aligned}$$

where $\left(\frac{a}{p}\right)$ is the Kronecker symbol. If we now add the point at infinity into our total we get $p-\left(\frac{a}{p}\right)$, so $a_p = p-(p-\left(\frac{a}{p}\right)) = \left(\frac{a}{p}\right) = \pm 1$. In this case we say that E has *multiplicative reduction* at p, and distinguish the cases $a_p = 1$ and $a_p = -1$ as *split* and *non-split* respectively. One can show that in the split case $E_p^{\mathrm{ns}}(\mathbb{F}_p)$ is isomorphic to the multiplicative group $\mathbb{F}_p^\times$, and in the non-split case it is isomorphic to the multiplicative subgroup of $\mathbb{F}_{p^2} = \mathbb{F}_p[x]/(x^2-a)$ consisting of the norm 1 elements; see [18, §2.10].

To sum up, there are three possibilities for $a_p = p - \#E_p^{\mathrm{ns}}(\mathbb{F}_p)$:

$$a_p = \begin{cases} 0 & \text{additive reduction,}\\ +1 & \text{split multiplicative reduction,}\\ -1 & \text{non-split multiplicative reduction.}\end{cases}$$

It can happen that the reduction type of E changes when we consider E as an elliptic curve over a finite extension $K/\mathbb{Q}$ (in which case we are then talking about reduction modulo primes $\mathfrak{p}$ of K lying above p). It turns out that this can only happen when E has additive reduction at p, which leads to the following definition.

Definition 24.30. An elliptic curve $E/\mathbb{Q}$ is *semistable* if it does not have additive reduction at any prime.

As we shall see in the next lecture, for the purposes of proving Fermat's Last Theorem, we can restrict our attention to semistable elliptic curves.

24.8 *L*-functions of elliptic curves versus *L*-functions of modular forms

Although we defined the L-function of an elliptic curve using an Euler product, we can always expand this product to obtain a Dirichlet series

$$L(E,s) = \prod_p \left(1-a_p p^{-s}+\chi(p)pp^{-2s}\right)^{-1} = \sum_{n=1}^{\infty} a_n n^{-s}.$$

We now observe that the integer coefficients a_n in the Dirichlet series for $L(E,s)$ satisfy the recurrence relations listed in (3) for an eigenform of weight $k=2$. We have $a_1 = 1$, $a_{mn} = a_m a_n$ for $m \perp n$, and $a_{p^{r+1}} = a_p a_{p^r} - p a_{p^{r-1}}$ for all primes p of good reduction, as you proved on Problem Set 7. For the primes of bad reduction we have $a_p \in \{0, \pm 1\}$ and it is easy to check that $a_{p^r} = a_p^r$, which applies to the coefficients of an eigenform in $S_k^{\mathrm{new}}(\Gamma_0(N))$ when $p|N$ (see Remark 24.16).

So it now makes sense to ask, given an elliptic curve $E/\mathbb{Q}$, is there a modular form f for which $L(E,s) = L(f,s)$? Or, to put it more simply, let $L(E,s) = \sum_{n=1}^{\infty} a_n n^{-s}$, and define

$$f_E(\tau) = \sum_{n=1}^{\infty} a_n q^n \qquad (q := e^{2\pi i \tau})$$

Our question then becomes: is $f_E(\tau)$ a modular form?

It's clear from the recurrence relation for a_{p^r} that if $f_E(\tau)$ is a modular form, then it must be a modular form of weight 2; but there are additional constraints. For $k = 2$ the equations (5) and (6) both give the Euler product

$$\prod_p \left(1 - a_p p^{-s} + \chi(p) p p^{-2s}\right)^{-1},$$

and it is essential that $\chi(p)$ is the same in both cases. For newforms $f \in S_k^{\text{new}}(\Gamma_0(N))$ we have $\chi(p) = 0$ for primes $p|N$, while for elliptic curves $E/\mathbb{Q}$ we have $\chi(p) = 0$ for primes $p|\Delta_{\min}(E)$. No elliptic curve over $\mathbb{Q}$ has good reduction at every prime, so we cannot use eigenforms of level 1, we need to consider newforms of some level $N > 1$.

This suggests we take N to be the product of the prime divisors of $\Delta_{\min}(E)$, but note that any N with the same set of prime divisors would have the same property, so this doesn't uniquely determine N. For semistable elliptic curves, it turns out that taking the product of the prime divisors of $\Delta_{\min}(E)$ is the correct choice, and this is all we need for the proof of Fermat's Last Theorem.

Definition 24.31. Let $E/\mathbb{Q}$ be a semistable elliptic curve. The *conductor* N_E of E is the product of the prime divisors of its minimal discriminant $\Delta_{\min}(E)$.

In general, the conductor N_E of an elliptic curve $E/\mathbb{Q}$ is always divisible by the product of the primes $p|\Delta_{\min}(E)$, and N_E is squarefree if and only if E is semistable. For primes p where E has multiplicative reduction (split or non-split) $p|N_E$ but $p^2 \nmid N_E$, and when E has additive reduction at p then $p^2|N_E$ and if $p > 3$ then $p^3 \nmid N_E$. The primes 2 and 3 require special treatment (as usual): the maximal power of 2 dividing N_E may be as large as 2^8, and the maximal power of 3 dividing N_E may be as large as 3^5, see [17, IV.10] for the details, which are slightly technical.

We can now say precisely what it means for an elliptic curve over $\mathbb{Q}$ to be modular.

Definition 24.32. An elliptic curve $E/\mathbb{Q}$ is *modular* if f_E is a modular form.

If $E/\mathbb{Q}$ is modular, the modular form f_E is necessarily a newform in $S_2^{\text{new}}(\Gamma_0(N_E))$ with an integral q-expansion; this follows from the Eichler-Shimura Theorem (see Theorem 24.37).

Theorem 24.33 (Modularity Theorem)**.** *Every elliptic curve $E/\mathbb{Q}$ is modular.*

Proof. This is proved in [4], which extends the results in [19, 20] to all elliptic curves $E/\mathbb{Q}$. □

Prior to its proof, the conjecture that every elliptic curve $E/\mathbb{Q}$ is modular was variously known as the Shimura-Taniyama-Weil conjecture, the Taniyama-Shimura-Weil conjecture, the Taniyama-Shimura conjecture, the Shimura-Taniyama conjecture, the Taniyama-Weil conjecture, or the Modularity Conjecture, depending on the author. Thankfully, everyone is now happy to call it the Modularity Theorem!

24.9 BSD and the parity conjecture

When E is modular, the L-function of E is necessarily the L-function of a modular form, and this implies that $L(E,s)$ has an analytic continuation and satisfies a functional equation, since this holds for the L-function of a modular form, by Theorem 24.25. Prior to the proof of the modularity theorem, this was an open question known as the Hasse-Weil conjecture; we record it here as a corollary to the Modularity Theorem.

Corollary 24.34. *Let E be an elliptic curve over $\mathbb{Q}$. Then $L(E,s)$ has an analytic continuation to a holomorphic function on $\mathbb{C}$, and the normalized L-function*

$$\tilde{L}_E(s) := N_E^{s/2}(2\pi)^{-s}\Gamma(s)L(E,s)$$

satisfies the functional equation

$$\tilde{L}_E(s) = w_E\tilde{L}_E(2-s),$$

where $w_E = \pm 1$.

The sign w_E in the functional equation is called the *root number* of E. If $w_E = -1$ then the functional equation implies that $\tilde{L}_E(s)$, and therefore $L(E,s)$, has a zero at $s=1$; in fact it is easy to show that $w_E = 1$ if and only if $L(E,s)$ has a zero of even order at $s=1$.

The conjecture of Birch and Swinnerton-Dyer (BSD) relates the order of vanishing of $L(E,s)$ at $s=1$ to the rank of $E(\mathbb{Q})$. Recall that

$$E(\mathbb{Q}) \simeq E(\mathbb{Q})_{\mathrm{tor}} \times \mathbb{Z}^r,$$

where $E(\mathbb{Q})_{\mathrm{tor}}$ denotes the torsion subgroup of $E(\mathbb{Q})$ and r is the *rank* of E.

Conjecture 24.35 (Weak BSD)**.** *Let $E/\mathbb{Q}$ be an elliptic curve of rank r. Then $L(E,s)$ has a zero of order r at $s=1$.*

The strong version of the BSD conjecture makes a more precise statement that expresses the leading coefficient of the Taylor expansion of $L(E,s)$ at $s=1$ in terms of various invariants of E. A proof of even the weak form of the BSD conjecture is enough to claim the Millennium Prize offered by the Clay Mathematics Institute. There is also the Parity Conjecture, which simply relates the root number w_E in the functional equation for $L(E,s)$ to the parity of r as implied by the BSD conjecture.

Conjecture 24.36 (Parity Conjecture)**.** *Let $E/\mathbb{Q}$ be an elliptic curve of rank r. Then the root number is given by $w_E = (-1)^r$.*

24.10 Modular elliptic curves

The relationship between elliptic curves and modular forms is remarkable and not at all obvious. It is reasonable to ask why people believed the modular conjecture in the first place. Probably the most compelling reason is that every newform of weight 2 with an integral q-series gives rise to an elliptic curve $E/\mathbb{Q}$.

Theorem 24.37 (Eichler-Shimura, Carayol)**.** *Let $f = \sum a_n q^n \in S_2^{\mathrm{new}}(\Gamma_0(N))$ be a newform with $a_n \in \mathbb{Z}$. There exists an elliptic curve $E/\mathbb{Q}$ of conductor N for which $f_E = f$.*

See [10, V.6] for an overview of how to construct the elliptic curve given by the theorem, which was known long before the modularity theorem was proved.[7] For a more detailed (but still very accessible) exposition, see [12].

The elliptic curve E whose existence is guaranteed by the Eichler-Shimura theorem is determined only up to isogeny.[8] This is due to the fact that isogenous elliptic curves E and E' over $\mathbb{Q}$ necessarily have the same L-function, which implies $f_E = f_{E'}$. If E and E' are isogenous over $\mathbb{Q}$ then their reductions modulo any prime p where they both have good reduction are necessarily isogenous, and as you showed on Problem Set 7, they must have the same trace of Frobenius a_p; it turns out that in fact E and E' must have the same reduction type at every prime so their L-functions are actually identical. The converse also holds; in fact, something even stronger is true; this follows from work begun by Tate and completed by Faltings in 1983 [7]; see [10, Thm. V.4.1] for further details.

Theorem 24.38 (Faltings-Tate)**.** *Let E and E' be elliptic curves over $\mathbb{Q}$ with L-function $L(E,s) = \sum a_n n^{-s}$ and $L(E',s) = \sum a'_n n^{-s}$, respectively. If $a_p = a'_p$ for sufficiently many primes p of good reduction for E and E', then E and E' are isogenous.*

What "sufficiently many" means depends on E and E', but it is a finite number. In particular, all but finitely many is always enough, which is all we need for the next lecture.

Corollary 24.39. *Elliptic curves $E, E'/\mathbb{Q}$ are isogenous if and only if $L(E,s) = L(E',s)$, equivalently, if and only if E_p and E'_p are isogenous modulo sufficiently many good primes p.*

The fact that isogenous elliptic curves have the same L-functions while distinct newforms have distinct L-functions means that the correspondence between elliptic curves and weight-2 newforms with $a_n \in \mathbb{Z}$ is many-to-one, not one-to-one; there can be up to 8 isomorphism classes of elliptic curves $E/\mathbb{Q}$ in the same isogeny class (but no more than 8, this is a result of Kenku [8]). But the modularity theorem implies that there is a one-to-one correspondence between isogeny classes of elliptic curves over $\mathbb{Q}$ and weight-2 newforms with $a_n \in \mathbb{Z}$.

For any given value of N, one can effectively enumerate the newforms in $S_2^{\text{new}}(\Gamma_0(N))$ with integral q-expansions; this is a finite list. It is also possible (but not easy)[9] to determine the isogeny classes of all elliptic curves of a given conductor N for suitable values of N, without assuming these elliptic curves are modular; this is also a finite list. When this was done for many small values of N, it was found that the two lists always matched perfectly. It was this matching that made the modularity conjecture truly compelling. Much of this matching was done before Theorems 24.37 and 24.38 had been completely proved, but they were both conjectured (and partially proved) much earlier.

References

[1] M. K. Agrawal, John H. Coates, David C. Hunt, Alfred J. van der Poorten, *Elliptic curves of conductor 11*, Math. Comp. **35** (1980), 991–1002.

[2] Amod Agashe, Kenneth Ribet, and William A. Stein, *The Manin constant*, Pure and Applied Mathematics Quarterly **2** (2006), 617–636.

[7] The original results of Eichler and Shimura [14] proved $a_p(E) = a_p(f)$ only for primes of good reduction and did not address the correspondence between the level and the conductor. The correspondence between the level and conductor was conjectured by Weil but not rigorously proved until 1986 by Carayol [5, §0.8].

[8] But there is an *optimal* representative for each isogeny class; see John Cremona's appendix to [2].

[9] This requires enumerating all solutions to certain Diophantine equations; see [1] and [11] for examples.

[3] A.O.L. Atkin and J. Lehner, *Hecke operators on $\Gamma_0(m)$*, Mathematische Annalen **185** (1970), 134–160.

[4] Christophe Breuil, Brian Conrad, Fred Diamond, and Richard Taylor, *On the modularity of elliptic curves over $\mathbb{Q}$: wild 3-adic exercises*, Journal of the AMS **14** (2001), 843–939.

[5] Henri Carayol, *Sur les représentations l-adiques associées aux formes modulaires de Hilbert*, Ann. Sci. École Norm. Sup. (4) **19** (1986), 409–468.

[6] Fred Diamond and Jerry Shurman, *A first course in modular forms*, Springer, 2005.

[7] Gerd Faltings, *Finiteness theorems for abelian varieties over number fields*, Inventiones **73** (1983), 349–366.

[8] M. A. Kenku, *On the number of $\mathbb{Q}$-isomorphism classes of elliptic curves in each $\mathbb{Q}$-isogeny class*, Journal of Number Theory **15** (1982), 199–202.

[9] Michael Laska, *An algorithm for finding a minimal Weierstrass equation for an elliptic curve*, Mathematics of Computation **38** (1982), 257-260.

[10] J. S. Milne, *Elliptic curves*, BookSurge Publishers, 2006.

[11] Andrew P. Ogg, *Abelian curves of small conductor*, J. Reine Angew. Math. **224** (1967), 204–215.

[12] Corentin Perent-Gentil, *Associating abelian varieties to weight-2 modular forms: the Eichler-Shimura construction*, Master's thesis, EPF Lausanne, 2014.

[13] Jean-Pierre Serre, *A course in arithmetic*, Springer, 1973.

[14] Goro Shimura, *Correspondances modulaires et les fonctions ζ de courbes algébriques*, Journal of the Mathematical Society of Japan, **10** (1958), 1–28.

[15] Goro Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan **11**, 1971.

[16] Joseph H. Silverman, *The arithmetic of elliptic curves*, second edition, Springer, 2009.

[17] Joseph H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Springer, 1994.

[18] Lawrence C. Washington, *Elliptic curves: Number theory and cryptography*, second edition, Chapman and Hall/CRC, 2008.

[19] Richard Taylor and Andrew Wiles, *Ring-theoretic properties of certain Hecke algebras*, Annals of Mathematics **141** (1995), 553–572.

[20] Andrew Wiles, *Modular elliptic curves and Fermat's last theorem*, Annals of Mathematics **141** (1995), 443-551.

18.783 / 18.7831 Elliptic Curves
Fall 2025