In this lecture we give a brief overview of *modular forms*, focusing on their relationship to elliptic curves. This connection is crucial to Wiles' proof of Fermat's Last Theorem [7]; the crux of his proof is that every *semistable* elliptic curve over $\mathbb{Q}$ is *modular*.[1] In order to explain what this means, we need 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 trivial proofs.

24.1 Modular forms

Definition 24.1. A holomorphic function $f\colon \mathbb{H} \to \mathbb{C}$ 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 $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$.

The j-function $j(\tau)$ is a weak modular form of weight 0 for $\Gamma_0(1) = \mathrm{SL}_2(\mathbb{Z})$, and $j(N\tau)$ is a weak modular form of weight 0 for $\Gamma_0(N)$. As an example of a weak modular form of positive weight, consider the Eisenstein series

$$G_k(\tau) = G_k([1,\tau]) = \sum{}' \frac{1}{(m+n\tau)^k},$$

which, for $k \geq 3$, is a weak modular form of weight k for $\Gamma_0(1)$. 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

$$G_k(S\tau) = G_k(-1/\tau) = \sum{}' \frac{1}{(m-\frac{n}{\tau})^k} = \sum{}' \frac{\tau^k}{(m\tau - n)^k} = \tau^k G_k(\tau),$$
$$G_k(T\tau) = G_k(\tau+1) = G_k(\tau) = 1^k G(\tau).$$

Note that if Γ contains $-I$, we must have $f(\tau) = (-1)^k f(\tau)$, which implies that the only weak modular form of odd weight for Γ is the zero function. We are specifically interested in the case $\Gamma = \Gamma_0(N)$, which does contain $-I$, thus we will restrict our attention to modular forms of even weight (some authors use $2k$ in place of k for precisely this reason).

As with modular functions (see Lecture 21), 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 $\tau \in \mathbb{H}$, since for the matrix T^N we have $c = 0$ and $d = 1$, so $(c\tau + d)^k = 1^k = 1$. It follows that $f(\tau)$ has a q-expansion of the form

$$f(\tau) = f^*(q^{1/N}) = \sum_{n=-\infty}^{\infty} a_n q^{n/N},$$

where $q = e^{2\pi i\tau}$. We say that f is *holomorphic at* ∞ if f^* is holomorphic at 0, equivalently, $a_n = 0$ for all $n < 0$. 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.

[1] We now know that every elliptic curve over $\mathbb{Q}$ is modular [1], whether it is semistable or not.

Andrew V. Sutherland

Definition 24.2. A *modular form* f is a weak modular form that is *holomorphic at the cusps*. Equivalently, f extends to a holomorphic function on the extended upper half plane $\mathbb{H}^* = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$.

The only modular forms of weight 0 are constant functions. This is main motivation for introducing the notion of weight, it allows us to generalize the notion of a modular function in an interesting way, by strengthening its analytic properties (it must be holomorphic, not just meromorphic) at the expense of weakening its 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 function $G_K(\tau)$ are modular forms. For $\Gamma_0(1)$ we have just one cusp orbit, so to show that $G_K(\tau)$ is holomorphic at the cusps we just need to check that

$$\lim_{\operatorname{im}\tau\to\infty} G_k(\tau) = \lim_{\operatorname{im}(\tau)\to\infty} \sum\nolimits' \frac{1}{(m+n\tau)^k} = 2\sum_{n=1}^{\infty} \frac{1}{n^k} = 2\zeta(k) < \infty,$$

which holds for all even $k \geq 4$ (recall that the series converges absolutely, which justifies rearranging the terms of the sum).

Definition 24.3. A modular form is called a *cusp* form if it vanishes at all the cusps. Equivalently, its q-expansion at every cusp has constant coefficient $a_0 = 0$

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, since

$$g_2(\infty) = 120\zeta(4) = \frac{4\pi^4}{3}, \qquad g_3(\infty) = 280\zeta(6) = \frac{8\pi^6}{27}, \qquad \Delta(\infty) = 0,$$

as shown in Lecture 18 (see Theorem 18.5).

Definition 24.4. The set of all modular forms of weight k for $\Gamma_0(N)$ is denoted $M_k(\Gamma_0(N))$. The subset of cusp forms in $M_k(\Gamma_0(N))$ is denoted $S_k(\Gamma_0(N))$.

It is clear that the sets $M_K(\Gamma_0(N))$ and $S_k(\Gamma_0(N))$ are both $\mathbb{C}$-vector spaces; in fact, they are finite dimensional vector spaces. The modular forms in $M_k(\Gamma_0(N))$ are said to be modular forms of *level* N

Example 24.5. Every modular form in $M_k(\Gamma_0(1))$ is a linear combination of products $G_4^a G_6^b$ where $4a + 6b = k$. The dimension of $M_k(\Gamma_0(1))$ is therefore equal to the number of solutions to $4a + 6b = k$ in non-negative integers. The dimension of $M_2(\Gamma_0(1))$ is zero, so there are no nonzero modular forms of weight 2 and level 1, and $M_k(\Gamma_0(1))$ is 1-dimensional for $k = 4, 6, 8, 10$. Asymptotically, the dimension of $M_k(\Gamma_0(1))$ approaches $k/12$.

For the vector space $S_k(\Gamma_0(N))$, there is a particular choice of basis that has some very nice properties. In order to define this basis, we need to introduce the *Hecke operators*. For each positive integer n, the Hecke operator $T(n)$ is a linear operator on the vector space $M_k(\Gamma_0(N))$ that fixes the subspace of cusp forms, so it is also a linear operator on $S_k(\Gamma_0(N))$. Our interest in the Hecke operators is that, if we normalize things appropriately, there is a unique basis for $S_k(\Gamma_0(N))$ whose elements are simultaneous eigenvectors (called *eigenforms*) for all of the Hecke operators.

24.2 Hecke operators

In order to motivate the definition of the Hecke operators on modular forms, we first define them in terms of lattices. For each positive integer n, the Hecke operator T_n sends a lattice $L = [\omega_1, \omega_2]$ to the formal sum of its index-n sublattices:

$$T_nL = \sum_{[L:L']=n} L' = \sum_{ad=n,\, 0\le b<d} [d\omega_1, a\omega_1 + b\omega_2]. \tag{1}$$

More formally, let $\mathcal{L}$ be the set of all (rank 2) lattices in the complex plane, and let $\mathrm{Div}(\mathcal{L})$ be the free abelian group generated by $\mathcal{L}$. Then Tn is the endomorphism of $\mathrm{Div}(\mathcal{L})$ determined by (2). Another important set of endomorphisms of $\mathrm{Div}(L)$ are the homethety operators R_λ defined by

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

for each $\lambda \in \mathbb{C}^*$. 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.

Theorem 24.6. *The operators T_n and R_λ satisfy the following:*

(i) $T_nR_\lambda = R_\lambda T(n)$ *and* $R_\lambda R_\mu = R_{\lambda\mu}$.

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

(iii) $T_{p^{n+1}} = T_{p^n}T_p - pT_{p^{n-1}}R_p$ *for all primes* p.

Moreover, the commutative algebra generated by the R_λ and the T_p contains all the T_n.

Proof. See [3, Prop. VII.5.1]. □

Remark 24.7. Recall that if $E/\mathbb{C}$ is the elliptic curve isomorphic to the torus $\mathbb{C}/L$, the index-n sublattices of L correpsond 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.

24.3 Hecke operators for modular forms of level 1

We now consider the action of the Hecke operators on modular forms for $\Gamma_0(1)$. The situation for modular forms of level $N > 1$ is entirely analogous, but the details are more complicated, so for the sake of simplicity we fix $N = 1$ throughout §24.3-24.5. We will address the issues involved in generalizing to $N > 1$ in §24.6

Recall that we originally define the Eisenstein series $G_k(L) = \sum' \omega^{-k}$ as a sum over the nonzero points ω in the lattice L, and then defined the function $G_k(\tau) = G([1,\tau])$ on the upper half plane. Thus we can view $G_k(L)$ as a function on lattices that satisfies $G_k(\lambda L) = \lambda^{-k}G_k(L)$.

Applying this perspective in reverse, we can view any modular function $f(\tau)$ as a function of the lattice $[1,\tau]$, and then extend this to arbitrary lattices $L = [\omega_1, \omega_2]$ by defining

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

where k is the weight of f and we order ω_1 and ω_2 so that ω_2/ω_1 is in the upper half plane. It then makes sense to define $R_\lambda f$ as

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

We define $T(n)f$ similarly, but introduce a scaling factor of n^{k-1} that will be convenient in what follows. Thus

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

It is a straight-forward exercise to verify that if f is a modular form of weight k and level 1, then so is $T(n)f$, and that $T(n)$ maps cusp forms to cusp forms. It is clear that $T(n)$ acts linearly, so it is a linear operator on the vector spaces $M_k(\Gamma_0(1))$ and $S_k(\Gamma_0(1))$. As an immediate consequence of Theorem 24.6, we have the following corollary.

Corollary 24.8. *$T_{mn} = T_mT_n$ for $m \perp n$ and $T_{p^{r+1}} = T_pT_{p^r} - p^{k-1}T_{p^{r-1}}$ for p prime.*

The corollary implies that it suffices to understand the behavior of T_p for p prime. Let us compute the the q-series expansion of T_pf, where $f(\tau) = \sum_{n=1}^\infty a_nq^n$ is a cusp form of weight k and level 1.

$$\begin{aligned}
(T_pf)(\tau) &= p^{k-1}\sum_{ad=p,\,0\le b<d} d^{-k} f\left(\frac{a\tau+b}{d}\right)\\
&= p^{k-1}f(p\tau) + p^{-1}\sum_{b=0}^{p-1} f\left(\frac{\tau+b}{p}\right)\\
&= p^{k-1}\sum_{n=1}^\infty a_nq^{pn} + p^{-1}\sum_{b=0}^{p-1}\sum_{n=1}^\infty a_n\zeta_p^{bn}q^{n/p}\\
&= p^{k-1}\sum_{n=1}^\infty a_{n/p}q^{n} + p^{-1}\sum_{n=1}^\infty a_n\left(\sum_{b=0}^{p-1}\zeta_p^{bn}\right)q^{n/p}\\
&= \sum_{n=1}^\infty \left(a_{pn} + p^{k-1}a_{n/p}\right)q^n
\end{aligned}$$

where $\zeta_p = e^{2\pi i/p}$ and $a_{n/p} = 0$ if p does not divide n. This calculation yields the following theorem and corollary.

Theorem 24.9. *Let $f \in S_k(\Gamma_0(1)$ have q-expansion $\sum_{n=1}^\infty a_nq^n$, and let $\sum_{n=1}^\infty b_nq^n$ be the q-expansion of T_pf, with p prime. Then*

$$b_n = \begin{cases} a_{pn} & \text{if } p\nmid n,\\ a_{pn} + p^{k-1}a_{n/p} & \text{if } p \mid n.\end{cases}$$

Corollary 24.10. *Let $f \in S_k(\Gamma_0(1)$ have q-expansion $\sum_{n=1}^\infty a_nq^n$, and let $\sum_{n=1}^\infty b_nq^n$ be the q-expansion of T_n. Then $b_1 = a_n$.*

Proof. This follows immediately from Theorem 24.9 and Corollary 24.8. □

24.4 Eigenforms of level 1

The Hecke operators T_n 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 of the Hecke operators; this means that $T_nf = \lambda_nf$ for some eigenvalue $\lambda_n \in \mathbb{C}^*$

of T_n, for all $n \geq 1$. When such an $f(\tau) = \sum_{n=1}^{\infty} a_n q^n$ has leading coefficient $a_1 = 1$, we call it an *eigenform*. Our goal is to construct a basis of eigenforms for $S_k(\Gamma_0(1))$, and to prove that it is unique. In order to do so, we need to introduce the *Peterson inner product*.

Definition 24.11. The *Peterson inner product* on $S_k(\Gamma_0(1))$ is defined by

$$\langle f, g \rangle = \int_{\mathcal{F}} f(\tau)\overline{g(\tau)} y^{k-2} dx dy, \tag{3}$$

where the integral ranges over points $\tau = x + yi$ in a fundamental region $\mathcal{F}$ for $\Gamma_0(1)$.

It is easy to check that $\langle f, g \rangle$ is a positive definite Hermitian form (it is a bilinear form that satisfies $\langle f, g \rangle = \overline{\langle g, f \rangle}$ and $\langle f, f \rangle \geq 0$ with equality only when $f = 0$), thus it defines an inner product for the complex vector space $S_k(\Gamma_0(1))$.

The Hecke operators are all self-adjoint with respect to the Peterson 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 by Corollary 24.8, they all commute with each other. This makes it possible to apply the following lemma.

Lemma 24.12. *Let V be a finite-dimensional $\mathbb{C}$-vector space equipped with a positive definite Hermition form, and let $\alpha_1, \alpha_2, \ldots$ be a sequence of commuting Hermitian operators. Then $V = \oplus V_i$, where each V_i is an eigenspace of every α_n.*

Proof. The matrix for α_1 is Hermitian, therefore diagonalizable, so we can decompose V as a direct sum of eigenspaces for α_1, writing $V = \oplus V(\lambda_i)$, 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. $\square$

So let us apply Lemma 24.12, and decompose $S_k(\Gamma_0(1)) = \oplus V_i$ as a direct sum of eigenspaces for the Hecke operators T_n. If $f(\tau) = \sum_{n=1}^{\infty} a_n q^n$ is a nonzero element of one of the V_i, then by Corollary 24.8, the coefficient b_1 in the q-expansion of $(T_n f)(\tau) = \sum_{n-1}^{\infty} b_n q^n$ is a_n. But we also have $T_n f = \lambda_n f$, for some eigenvalue λ_n of T_n, and therefore $a_n = \lambda_n a_1$. This implies $a_1 \neq 0$, since otherwise $f = 0$, and if we normalize f so that $a_1 = 1$, we have $a_n = \lambda_n$ for all $n \geq 1$, and f is then uniquely determined by the sequence of Hecke eigenvalues λ_n for V_i. It follows that each V_i is one-dimensional and contains element with $a_1 = 1$, that is, an eigenform. We record this result in the following theorem.

Theorem 24.13. *The vector space $S_k(\Gamma_0(1))$ can be written as a direct sum of 1-dimensional eigenspaces for the Hecke operators T_n and has a unique basis of eigenforms $f(\tau) = \sum_{n=1}^{\infty} a_n q^n$, where each a_n is the eigenvalue of T_n on the 1-dimensional subspace generated by f.*

Corollary 24.14. *Let $f(\tau) = \sum_{n=1}^{\infty} a_n q^n$ be an eigenform in $S_k(\Gamma_0(1)$. Then $a_{mn} = a_m a_n$ for all $m \perp n$, and $a_{p^{r+1}} = a_p a_{p^r} - p^{k-1} a_{p^{r-1}}$ for all primes p.*

In the case $k = 2$, the prime-power recurrence in 24.14 should look familiar — it is exactly the same as the recurrence satisfied by the Frobenius traces $a_{p^r} = p^r + 1 - \#E(\mathbb{F}_{p^r})$ of an elliptic curve $E/\mathbb{F}_p$, which you proved in Problem Set 7.

24.5 L-series associated to modular forms

Our interest in cusp forms is that there is an *L-series* associated to each cusp form.

Definition 24.15. The *L-series* of a cusp form $f(\tau) = \sum_{n=1}^{\infty} a_n q^n$ is the function

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

which converges uniformly for $\mathrm{Re}(s) > 1 + k/2$, where k is the weight of f.

The function $L_f(s)$ is an example of a *Dirichlet L-series.* Before examining its properties, we first recall some general facts about Dirichlet series and Dirichlet L-series.

Definition 24.16. A *Dirichlet series* is a complex function of the form $f(s) = \sum_{n=1}^{\infty} a_n n^{-s}$.
A *Dirichlet L-series* is a Dirichlet series of the form

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

where χ is a *Dirichlet character*, a completely multiplicative function $\chi\colon \mathbb{Z} \to \mathbb{C}$ that restricts to a group character on $(\mathbb{Z}/m\mathbb{Z})^*$, for some positive integer m for which $\chi(n) \neq 0 \Leftrightarrow n \perp m$ (when $m = 1$ then $\chi(n) = 1$ is the trivial character). This series converges for $\mathrm{re}\, s > 1$ and can be analytically continued to a meromorphic function on $\mathbb{C}$.

Example 24.17. The *Riemann zeta function*

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

is the Dirichlet L-series for the trivial character. It's analytic continuation is holomorphic everywhere except at $s = 1$, where it has a simple pole.

The following theorems illustrate two key properties of Dirichlet L-series in the particular case of $\zeta(s)$. The first is the existence of an Euler product.

Theorem 24.18. *For* $\mathrm{Re}(s) > 1$ *we have*

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1},$$

where the product is over primes.

Proof. Since $\zeta(s)$ converges absolutely for $\mathrm{Re}(s) > 1$, we have

$$\prod_p (1 - p^{-s})^{-1} = \prod_p (1 + p^{-s} + p^{-2s} + \dots) = \sum_{n=1}^{\infty} n^{-s} = \zeta(s). \qquad \square$$

The Euler product for a general Dirichlet L-series $L(x, \chi)$ is

$$L(s, \chi) = \prod_p (1 - \chi(p) p^{-s})^{-1} \qquad (\mathrm{Re}(s) > 1).$$

The second key property of a Dirichlet L-series is its functional equation.

Theorem 24.19. *Let*

$$\tilde{\zeta}(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s),$$

where $\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}dt$ *is the gamma function. Then*

$$\tilde{\zeta}(s) = \tilde{\zeta}(1-s).$$

We can think of the function $\tilde{\zeta}(s)$ as a "normalized" $\zeta(s)$; different Dirichlet L-series have different normalization factors, but once they are suitably normalized they all satisfy a functional equation similar to the one given above, with an evaluation at s on one side and an evaluation at $1-s$ on the other.

Returning to our discussion of modular forms, the L-series $L_f(s)$ of a cusp form f for $\Gamma_0(1)$ also satisfies a functional equation.

Theorem 24.20. *Let* $f \in S_k(\Gamma_0(1))$ *be a cusp form with* L*-series* $L_f(s)$*. Then* $L_f(s)$ *extends analytically to a holomorphic function on* $\mathbb{C}$*, and*

$$\tilde{L}_f(s) = (2\pi)^{-s}\Gamma(s)L_f(s).$$

satisfies the functional equation

$$\tilde{L}_f(s) = (-1)^{k/2}\tilde{L}_f(k-s).$$

In the case that f is an eigenform, we get an Euler product for $L_f(s)$. This is not true for arbitrary cusp forms, and as we shall see shortly, in order to relate elliptic curves to modular forms, the existence of an Euler product is crucial.

Theorem 24.21. *Let* T_n *denote the* n*th Hecke operator on* $S_k(\Gamma_0(1))$*. Then*

$$\sum_{n=1}^{\infty} T_n n^{-s} = \prod_p (1 - T_p p^{-s} + p^{k-1}p^{-2s})^{-1},$$

and if $f \in S_k(\Gamma_0(1))$ *is an eigenform with* L*-series* $L_f(s)$*, we have the Euler product*

$$L_f(s) = \sum_{n=1}^{\infty} a_n n^{-s} = \prod_p (1 - a_p p^{-s} + p^{k-1}p^{-2s})^{-1}.$$

24.6 Eigenforms of level *N*.

So far we have dealt only with cusp forms of level 1. Everything we have seen can be generalized to arbitrary level N, but there are two issues that arise when doing so.

The first issue is that when considering cusp forms for $\Gamma_0(N)$, we really want to restrict our attention to cusp forms that are "new" at level N, meaning that they are not also cusp forms for $\Gamma_0(d)$, for some $d|N$. The reason for this is that while it is still true that $S_k(\Gamma_0(N))$ is spanned by eigenforms of the Hecke operators, the eigenspaces will not be 1-dimensional unless we restrict to the subspace of newforms.

To define the subspace of new forms, we first deal with the "old" forms'. Let $S_k^{old}(\Gamma_0(N))$ be the subspace spanned by $\bigcup S_k(N')$ where N' ranges over all N' properly dividing N. Now let $S_k^{new}(\Gamma_0(N))$ be the subspace orthogonal to $S_k^{old}(\Gamma_0(N))$ in $S_k(\Gamma_0(N))$. The Hecke eigenspaces of $S_k^{new}(\Gamma_0(N))$ are then 1-dimensional, and each eigenspace is generated by a uniquely determined (normalized) eigenform that we call a *newform*.[2]

The second issue is that the primes p that divide N require special attention. To deal with this, we let χ be the trivial character for $(\mathbb{Z}/N\mathbb{Z})^*$; that is, $\chi(m) = 1$ if $\gcd(m, N) = 1$, and $\chi(m) = 0$ otherwise. Then the Euler product identity for a newform in $S_k^{\rm new}(\Gamma_0(N))$ is

$$L_f(s) = \prod_p (1 - a_p p^{-s} + \chi(p)p^{k-1}p^{-2s})^{-1}. \tag{4}$$

24.7 The *L*-series 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-series with an Euler product. In fact, with elliptic curves, we use the Euler product to define the L-series.

Definition 24.22. The *L-series of an elliptic curve* $E/\mathbb{Q}$ is

$$L_E(s) = \prod_p L_p(p^{-s})^{-1} = \prod_p \left(1 - a_p p^{-s} + \chi(p)pp^{-2s}\right)^{-1}, \tag{5}$$

where the Dirichlet character $\chi(p)$ is 0 if E has bad reduction at p, and 1 otherwise.[3] For primes p where E has good reduction (all but finitely many), a_p is the Frobenius trace $p + 1 - \#E_p(\mathbb{F}_p)$, where E_p is the reduction of E modulo p. Equivalently, the polynomial $L_p(T)$ is the numerator of the zeta function

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

that appeared in Problem Set 7. For primes p where E has bad reduction, the polynomial $L_p(T)$ is defined by

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

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

The L-series $L_E(s)$ converges for $\Re(s) > 3/2$. As we will see shortly, the question of whether or not $L_E(s)$ has an analytic continuation is intimately related to the question of modularity (we now know the answer is yes, since every elliptic curve over $\mathbb{Q}$ is modular).

24.8 Determining 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

[2] In the interest of full disclosure, we should note that the formulas for the action of the Hecke operators become rather more complicated for level $N > 1$, but this does not concern us here; all we need to know is that they exist and satisfy Corollary 24.8.

[3] As explained in §24.8, this assumes we are using a minimal Weierstrass equation for E.

that if $y^2 = x^3 + Ax + B$ is a Weierstrass equation for E, then, up to isomorphism, so is $y^2 + u^4Ax + u^6B$, 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 isomorphic equation 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 $y + 1/2$).

Definition 24.23. Let $E/\mathbb{Q}$ be an elliptic curve. A *minimal Weierstrass equation* for E is a general Weierstrass equation

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

with $a_1, a_2, a_3, a_4, a_6 \in \mathbb{Z}$ that defines an elliptic curve $E'/\mathbb{Q}$ that is isomorphic to E over $\mathbb{Q}$ whose discriminant $\Delta(E')$ divides the discriminant of every other such elliptic curve. The discriminant $\Delta(E')$ is called the *minimal discriminant* of E and is denoted $\Delta_{\min}(E)$.

It is not immediately obvious that an elliptic curve necessarily has a minimal Weierstrass equation, but for elliptic curves over $\mathbb{Q}$ this is indeed the case; see [4, Prop. VII.1.3]. It can be computed in Sage via `E.minimal_model()`; see [2] for algorithm details.

We now address the three cases of bad reduction. To simplify matters, we will ignore the prime 2. At any odd prime p of bad reduction we can represent $E_p/\mathbb{F}_p$ by an equation of the form $y^2 = f(x)$, for some cubic $f \in \mathbb{F}_p[x]$ that has a repeated root. We can choose $f(x)$ so that this repeated root is at 0, and it is easy to verify that there is then exactly on singular point of E_p, which occurs at the affine point $(0, 0)$.

If we exclude the point $(0, 0)$, the standard algebraic formulas for the group law on $E(\mathbb{F}_p)$ still work, and the set

$$E_p^{\text{ns}}(\mathbb{F}_p) = E_p(\mathbb{F}_p)\backslash\{0, 0\}$$

of non-singular points of $E_p(\mathbb{F}_p)$ is actually closed under the group operation. Thus $E_p^{\text{ns}}(\mathbb{F}_p)$ is a finite abelian group, and we define

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

This is completely analogous to the nonsingular case, where $a_p = p + 1 - \#E(\mathbb{F}_p)$; we have removed the point $(0, 0)$ from consideration, so we should "expect" the cardinality of $E_p^{\text{ns}}(\mathbb{F}_p)$ to be p, rather than $p + 1$, and a_p measures the deviation from this value.

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

- **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 normalize the points so that $y = 1$, 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 on our original curve). It follows that $E_p^{\text{ns}}(\mathbb{F}_p)$ is a cyclic group of order p, which is isomorphic to the additive group of $\mathbb{F}_p$; see [6, §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 $zy^2 = x^3 + ax^2 z$, and the point $(0 : 1 : 0)$ at infinity is the only non-singular point on the curve whose x-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 normalize $z = 1$. This yields the affine curve $t^2 = x + a$, and the number of points with $x \neq 0$ is

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

  where $\left(\frac{a}{p}\right)$ is the Kronecker symbol. If we now add the point at infinity back in we get a total of $p - \left(\frac{a}{p}\right)$ points, thus $a_p = \left(\frac{a}{p}\right)$.

  In this case we say that E has *multiplicative reduction* at p, and further distinguish the cases $a_p = 1$ and $a_p = -1$ as *split* and *non-split* respectively. One can show that in the former case $E_p^{\mathrm{ns}}(\mathbb{F}_p)$ is isomorphic to the multiplicative group $\mathbb{F}_p^*$, and in the latter case it is isomorphic to the multiplicative subgroup of $\mathbb{F}_{p^2} = \mathbb{F}_p[x]/(x^2 - a)$ made up by the elements of norm 1; see [6, §2.10].

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

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

There is one further issue to consider. It could happen that the reduction type of E at a prime p changes when we consider E as an elliptic curve over an extension of $\mathbb{Q}$ (this gives us more flexibility when looking for a minimal Weierstrass equation). It turns out that this can only happen when E has additive reduction at p. This leads to the following definition.

Definition 24.24. An elliptic curve $E/\mathbb{Q}$ is *semi-stable* if it does not have additive reduction at any prime.

As we shall see, for the purposes of proving Fermat's Last Theorem, we can restrict our attention to semi-stable elliptic curves.

24.9 *L*-series of elliptic curves and *L*-series of modular forms

Having fully defined the L-series $L_E(s) = \prod_p (L_p(p^{-s})^{-1} = \sum_{n=1}^{\infty} a_n n^{-s}$ of an elliptic curve $E/\mathbb{Q}$, we now note that the coefficients a_n satisfy all the relations satisfied by the coefficients of a weight-2 eigenform. We have $a_1 = 1$, and, as in Corollary 24.8, we have $a_{mn} = a_m a_n$ for all $m \perp n$, and $a_{p^{r+1}} = a_p a_{p^r} - p^{k-1} a_{p^{r-1}}$ for all primes p, with $k = 2$.

So now we might 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

$$f_E(\tau) = \sum_{n=1}^{\infty} a_n q^n.$$

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 (4) and (5) 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 the Dirichlet character χ is the same in both cases. 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, in which case χ is the trivial character for $(\mathbb{Z}/N\mathbb{Z})^*$.

For $L_E(s)$ we know that $\chi(p) = 0$ if and only if p divides $\Delta_{\min}(E)$. This suggests taking N to be the product of the prime divisors of $\Delta_{\min}(E)$, but we should note that any N with the same set of prime divisors would have the same property. In turns out that for semi-stable elliptic curves, simply taking the product of the prime divisors of $\Delta_{\min}(E)$ is the right thing to do, and this is all we need for the proof of Fermat's Last Theorem.

Definition 24.25. Let $E/\mathbb{Q}$ be a semi-stable elliptic curve with minimal discriminant $\Delta_{\min}$. The *conductor* N_E of E is the product of the prime divisors of $\Delta_{\min}$.

Remark 24.26. For elliptic curves that are not semistable, at primes $p > 3$ where E has additive reduction we simply replace the factor p in N_E by p^2. But the primes 2 and 3 require special treatment (as usual), and the details can get quite technical; see [5, IV.10]. In any case, the conductor of an elliptic curve $E/\mathbb{Q}$ is squarefree if and only if it is semistable.

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

Definition 24.27. $E/\mathbb{Q}$ is *modular* if $f_E(\tau)$ is a modular form of weight 2 for $\Gamma_0(N_E)$.

References

[1] 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 American Mathematical Society **14** (2001), 843–939.

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

[3] J.-P. Serre, *A course in arithmetic*, Springer, 1973.

[4] Joseph H. Silverman, *The arithmetic of elliptic curves*, 2nd edition, Springer, 2009.

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

[6] Lawrence C. Washington, *Elliptic curves: Number theory and cryptography*, 2nd edition, CRC Press, 2008.

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

MIT OpenCourseWare
http://ocw.mit.edu

18.783 Elliptic Curves
Spring 2013

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.