12
Modular Forms and Elliptic Curves

Let E be an elliptic curve over $\mathbb{Q}$, so E is defined by an equation $y^2 = x^3 + ax + b$ with $a, b \in \mathbb{Q}$. It was recently proved by Andrew Wiles and others that all such elliptic curves are "modular", a result which provides a huge number of tools for studying elliptic curves over $\mathbb{Q}$. Two important consequences are that Fermat's Last Theorem is true, and that the conjecture of Birch and Swinnerton-Dyer about the rank of $E(\mathbb{Q})$ (see Chapter 14) involves objects that are defined.

In Section 12.1 we define modular forms, and in Section 12.2 we give a definition of what it means for an elliptic curve to be modular. Section 12.3 contains a brief discussion of how modularity of elliptic curves implies the truth of Fermat's Last Theorem.

12.1 Modular Forms

The complex *upper half plane* is the set

$$\mathfrak{h} = \{z \in \mathbb{C} : \mathrm{Im}(z) > 0\},$$

and the *upper half plane union the cusps* is

$$\mathfrak{h}^* = \mathfrak{h} \cup \mathbb{Q} \cup \{\infty\},$$

where ∞ should be viewed as being at infinity along the i axis. A *holomorphic function* $f : \mathfrak{h} \to \mathbb{C}$ is a function such that for all $z \in \mathfrak{h}$ the derivative

$$f'(z) = \lim_{h \to 0} \frac{f(z+h) - f(z)}{h}$$

exists. Holomorphicity is a very strong condition because $h \in \mathbb{C}$ can approach 0 in many ways. Let $\mathcal{H}$ denote the complex vector space of holomorphic functions on $\mathfrak{h}$.

Recall that $\mathrm{SL}_2(\mathbb{Z})$ denotes the group of 2×2 integer matrices with determinant 1. The linear fractional transformation induced by $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$ is

$$\gamma(z) = \frac{az+b}{cz+d}.$$

The group $\mathrm{SL}_2(\mathbb{Z})$ acts on the right on $\mathcal{H}$ by pre-composition:

$$f(z) \mapsto f(\gamma(z)).$$

The space of *holomorphic differentials on* $\mathfrak{h}$ is the complex vector space of expressions

$$\Omega = \{f(z)dz : f \text{ is a holomorphic function on } \mathfrak{h}\}.$$

There is a bijection between the holomorphic functions on $\mathfrak{h}$ and the holomorphic differentials on $\mathfrak{h}$ given by $f(z) \mapsto f(z)dz$. The group $\mathrm{SL}_2(\mathbb{Z})$ acts on Ω in a different and more interesting way than it acts on $\mathcal{H}$. For $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ and $f(z)dz \in \Omega$, let

$$(f(z)dz)|_\gamma = f(\gamma(z))d(\gamma(z)).$$

Remark 12.1.1. The quotient rule from calculus and that $\det(\gamma) = 1$ imply that $d(\gamma(z)) = (cz+d)^{-2}dz$. Thus under the bijection between Ω and $\mathcal{H}$, the action of $\mathrm{SL}_2(\mathbb{Z})$ on Ω corresponds to the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathcal{H}$ given by

$$f(z)|_\gamma = f(\gamma(z))(cz+d)^{-2}.$$

For any positive integer N, consider the subgroup

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \,:\, N \mid c \right\} \subset \mathrm{SL}_2(\mathbb{Z}).$$

Let $\Omega(\Gamma_0(N))$ be the subspace of Ω of function f such that $f(z)dz$ is fixed by every element of $\Gamma_0(N)$. If $f(z)dz \in \Omega(\Gamma_0(N))$, then since $\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_0(N)$ for some integer $h > 0$ (in fact, $h = 1$), we have $f(z+h)dz = f(z)dz$, so $f(z+h) = f(z)$.

Proposition 12.1.2. *The holomorphic function $q(z) = e^{2\pi i z/h}$ maps the vertical strip*

$$V = \{z \in \mathfrak{h} : 0 \le |z| < h\}$$

bijectively onto the punctured open unit disk $D = \{z \in \mathbb{C} : 0 < |z| < 1\}$. If $f : \mathfrak{h} \to \mathbb{C}$ is a function that satisfies $f(z+h) = f(z)$, then there is a unique function $F : D \to \mathbb{C}$ such that $f(z) = F(q(z))$.

Proof. We will not prove that e is holomorphic, because this is a standard part of any complex analysis course. If $z = x + iy \in V$, then

$$e^{2\pi iz/h} = e^{2\pi i(x+iy)/h} = e^{-2\pi y/h} e^{2\pi ix/h}$$

is in D since $y > 0$. Every element of D is uniquely of the form $e^{-2\pi y/h} e^{2\pi ix/h}$ for $y/h > 0$ and $0 \leq x/h < 1$, so q is a bijection.

For $w \in D$ let $F(w) = f(q^{-1}(w))$, where $q^{-1} : D \to V$ is the inverse of q. Then for $z \in V$, we have $F(q(z)) = f(q^{-1}(q(z))) = f(z)$ as required. Since $f(z) = f(z+1)$, we have $F(q(z)) = f(z)$ for all $z \in \mathfrak{h}$. □

A differential $f(z)dz \in \Omega(\Gamma_0(N))$ is *holomorphic at infinity* if the function $F(q)$ of Proposition 12.1.2 on $D \subset \mathbb{C}$ extends to a holomorphic function at 0. If this extension (which is necessarily unique) is 0 at 0, we say that f *vanishes at infinity.* If f is holomorphic at infinity, then $F(q)$ has a Taylor expansion (it is a theorem in complex analysis that holomorphic functions converge to their Taylor expansions) and there is a neighborhood of infinity such that

$$f = \sum_{n=0}^{\infty} a_n q^n$$

for complex numbers a_n. This expansion is called the *q-expansion of $f(z)$ at infinity.*

The vector space of *modular forms (of weight 2) for* $\Gamma_0(N)$ is the space $M_2(\Gamma_0(N))$ of functions $f : \mathfrak{h} \to \mathbb{C}$ such that

- $f(z)dz \in \Omega(\Gamma_0(N))$
- For every $\alpha \in \mathrm{SL}_2(\mathbb{Z})$, the differential $(f(z)dz)|_\alpha$ is holomorphic at infinity.

We pause to comment on why the second condition makes sense. First note that if $\gamma \in \Gamma_0(N)$ and $\omega = f(z)dz$, then

$$(\omega|_\alpha)|_{\alpha^{-1}\gamma\alpha} = \omega|_{\gamma\alpha} = \omega|_\alpha.$$

If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and $t = \left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right)$ for $h \in \mathbb{Z}$, then there is an integer h such that

$$\delta = \alpha t \alpha^{-1} = \begin{pmatrix} -ach+1 & ha^2 \\ -hc^2 & hac+1 \end{pmatrix} \in \Gamma_0(N).$$

Thus $t = \alpha^{-1}\delta\alpha = \left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right)$, so $(\omega|_\alpha)|_t = \omega$ and $\omega|_\alpha$ has a Fourier expansion at ∞.

The subspace $S_2(\Gamma_0(N))$ of *cusp forms* is the subspace of elements $f \in M_2(\Gamma_0(N))$ such that $f(\gamma(z))$ vanishes at infinity for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$.

Remark 12.1.3. The condition the $f(\gamma(z))$ has a nice property at infinity for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ probably seems ad hoc. It is motivated by the following geometric observation. The quotient of $\mathfrak{h}$ by the action of $\Gamma_0(N)$ is a noncompact Riemann surface $Y_0(N)$ (it is missing a finite set set of points). Elements of $\Omega(\Gamma_0(N))$ correspond to differentials on $Y_0(N)$. Differentials on noncompact Riemann surfaces are not well behaved; for example, the space of holomorphic differentials will not be finite dimensional.

Suppose $f(z) \in M_2(\Gamma_0(N))$. We have $dq/q = dz$, so if $f(q) = \sum_{n=0}^{\infty} a_n q^n$, then the differential

$$f(z)dz = f(q)dq/q = \sum a_n q^{n-1} dq$$

is holomorphic at infinity if and only if $a_0 = 0$. This illustrates that the differentials on $Y_0(N)$ that extend to holomorphic differentials on the compactification $X_0(N)$ are exactly the elements of $S_2(\Gamma_0(N))$.

12.1.1 The Dimension of $S_2(\Gamma_0(N))$

Using techniques from algebraic geometry or complex analysis, one can prove that $S_2(\Gamma_0(N))$ has finite dimension, and more precisely

$$\dim S_2(\Gamma_0(N)) = 1 + \frac{\mu}{12} - \frac{\nu_2}{4} - \frac{\nu_3}{3} - \frac{\nu_\infty}{2},$$

where $\mu = N\prod_{p|N}(1+1/p)$, and $\nu_2 = \prod_{p|N}\left(1+\left(\frac{-4}{p}\right)\right)$ unless $4 \mid N$ in which case $\nu_2 = 0$, and $\nu_3 = \prod_{p|N}\left(1+\left(\frac{-3}{p}\right)\right)$ unless $2 \mid N$ or $9 \mid N$ in which case $\nu_3 = 0$, and $\nu_\infty = \sum_{d|N}\varphi(\gcd(d, N/d))$.

For example,

$$\dim_{\mathbb{C}} S_2(\Gamma_0(2)) = 1 + \frac{3}{12} - \frac{1}{4} - \frac{0}{3} - \frac{2}{2} = 0,$$

and

$$\dim_{\mathbb{C}} S_2(\Gamma_0(11)) = 1 + \frac{12}{12} - \frac{0}{4} - \frac{0}{3} - \frac{2}{2} = 1.$$

One can prove that the vector space $S_2(\Gamma_0(11))$ has basis

$$f = q\prod_{n=1}^{\infty}(1-q^n)^2(1-q^{11n})^2 = q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 - 2q^7 + \cdots.$$

Modular forms are a huge area. See [Ser73, Ch. 7] for an excellent first introduction then look at the modern survey paper [DI95] for an excellent overview of the most important basic facts about modular forms along with a fairly complete bibliography.