LECTURE 19, WEDNESDAY 14.04.04

FRANZ LEMMERMEYER

1. Outline

This time I will give you an idea of the set-up of Mazur's classification of torsion groups of elliptic curves defined over $\mathbb{Q}$. The big tool is a family of compact Riemann surfcaces $X_1(N)$; compact Riemann surfaces correspond to compex algebraic curves, and the points on these curves (except for a few points called cusps) correspond to isomorphism classes of elliptic curves with a point of order N; its rational points (except for some cusps) correspond to isomorphism classes of elliptic curves **defined over** $\mathbb{Q}$ with a **rational point of order** N. What Mazur did was prove that the only rational points on these curves $X_1(N)$, $N > 12$, are necessarily cusps. Sounds simple, but ...

What I'll do below is explain

- how to set up this bijection between isomorphism classes of elliptic curves with torsion points on the one hand, and points on some set on the other hand;
- how to make this set into a Riemann surface $Y_1(N)$;
- how to make $Y_1(N)$ into a compact Riemann surface $X_1(N)$ by "adding" cusps;
- how to realize the function field of $X_1(N)$ using modular functions;
- how to find an equation for the algebraic curve corresponding to the Riemann surface $X_1(N)$.

This is material for a whole semester or more, and I haven't even mentioned Hecke operators, Eichler-Shimura theory, or the Taniyama-Shimura conjecture (now a theorem due to Wiles et al.); here are a few books that will help you understand the basics:

(1) J.P. Serre, *A course in arithmetic*, Springer Verlag. Gives the basic theory of modular forms with respect to $\mathrm{SL}_2(\mathbb{Z})$ with Serre's usual clarity.
(2) A. Knapp, *Elliptic curves*, Princeton UP. Also very down to earth.
(3) G. Shimura, *Introduction to the arithmetic theory of automorphic forms*, Princeton UP. This book contains the details of the constructions, but some claim it is not very user-friendly.

2. Isomorphism Classes of Elliptic Curves

2.1. **Isomorphisms and Isogenies.** Let $\mathbb{H} = \{x + iy \in \mathbb{C} : y > 0\}$ denote the upper half plane. Let $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ denote a 2×2-matrix with real coefficients, and put $M(z) = \frac{az+b}{cz+d}$. The identity $\operatorname{Im} M(z) = \det M |cz + d|^{-2} \operatorname{Im} z$ shows that the group $\mathrm{GL}_2^+(\mathbb{R})$ of matrices with positive determinant act on $\mathbb{H}$. In particular, $\mathrm{GL}_2^+(\mathbb{Q})$ and $\mathrm{SL}_2(\mathbb{Z})$ act on the upper half plane.

Now consider lattices Λ, Λ' in $\mathbb{C}$ and the corresponding tori $E = \mathbb{C}/\Lambda$ and $E' = \mathbb{C}/\Lambda'$. We say that E and E' are isomorphic if there is a $\lambda \in \mathbb{C}^\times$ with $\lambda\Lambda = \Lambda'$ (note that this induces a map $E \longrightarrow E'$ sending $z + \Lambda$ to $\lambda z + \Lambda'$). We say that a map $\phi : E \longrightarrow E'$ is an isogeny if it is given by $z + \Lambda \longmapsto \lambda z + \Lambda'$, where $\lambda \in \mathbb{C}$ satisfies $\lambda\Lambda \subset \Lambda'$.

Example. If $\Lambda = \mathbb{Z} \oplus i\mathbb{Z}$, then multiplication by i induces an isomorphism, and multiplication by $1 + i$ induces an isogeny $E \longrightarrow E$ since $(1 + i)\Lambda \subset \Lambda$. Moreover, multiplication by m is always an isogeny $[m] : E \longrightarrow E$. The last example also shows that describing isogenies by equations on the elliptic curves side is much more involved than on the side of lattices.

By the way, the degree of an isogeny is the cardinality of the kernel. For example, multiplication by N is a map $[N]$ with kernel $E[N] \simeq \mathbb{Z}/N \oplus \mathbb{Z}/N$, hence $[N]$ has degree N^2.

Given any lattice, the map

$$\Lambda \longrightarrow \Lambda : m_1\omega_1 + m_2\omega_2 \longmapsto 2m_1\omega_1 + m_2\omega_2$$

defines an isogeny of degree 2, that is, a 2-isogeny: its kernel consists of the images of 0 and $\frac{1}{2}\omega_1$, corresponding to the point at infinity and some point of order 2 on $E = \mathbb{C}/\Lambda$. The isogeny $E \longrightarrow E$ will be defined over $\mathbb{Q}$ (i.e. the corresponding rational maps will have rational coefficients) if and only if the point corresponding to $\frac{1}{2}\omega_1$ has rational coordinates. Such 2-isogenies will be central in our proof of the Mordell-Weil theorem, although we will construct them in a completely different way.

2.2. **Isomorphism classes.** In the following, for any $\tau \in \mathbb{H}$ put $\Lambda_\tau = \mathbb{Z} \oplus \tau\mathbb{Z}$ and $E_\tau = \mathbb{C}/\Lambda_\tau$. A simple calculation shows

Proposition 1. *For $\tau, \tau' \in \mathbb{H}$,*

1. *E_τ and $E_{\tau'}$ are isomorphic if and only if there is an $M \in \mathrm{SL}_2(\mathbb{Z})$ such that $M(\tau) = \tau'$.*
2. *E_τ and $E_{\tau'}$ are isogenous if and only if there is an $M \in \mathrm{GL}_2^+(\mathbb{Q})$ such that $M(\tau) = \tau'$.*

The first result shows that there is a bijection between the isomorphism classes of complex elliptic curves and orbits of elements in the upper half plane under the action of $\mathrm{SL}_2(\mathbb{Z})$. This set of orbits is traditionally denoted by $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. The notation stems from the fact that $\mathrm{SL}_2(\mathbb{Z})$ acts from the left on $\mathbb{H}$. The points in $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$ are equivalence classes of elements $\tau \in \mathbb{H}$, where $\tau \sim \tau'$ if $M\tau = \tau'$ for some $M \in \mathrm{SL}_2(\mathbb{Z})$.

Note that this bijection is explicit: given some $\tau \in \mathbb{H}$, form the lattice Λ_τ and compute the associated coefficients g_2 and g_3; the action of $\mathrm{SL}_2(\mathbb{Z})$ on Λ_τ and hence on E_τ induces an admissible transformation $E \longrightarrow E'$.

The next step is adding structure to this picture: isomorphism classes of elliptic curves over $\mathbb{C}$ correspond to points in the space $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. The first thing to do is to find a fundamental domain for the action of $\mathrm{SL}_2(\mathbb{Z})$ on $\mathbb{H}$. Call two points $\tau, \tau' \in \mathbb{H}$ equivalent if $M(\tau) = \tau'$ for some $M \in \mathrm{SL}_2(\mathbb{Z})$. Then it is relatively easy to show that every $\tau \in \mathbb{H}$ is equivalent to some τ in the region defined by $|\tau| \geq 1$ and $-\frac{1}{2} \leq \mathrm{Re}\ \tau \leq \frac{1}{2}$, and that two points in that region can be equivalent only if they lie on the boundary. More exactly, the points $-\frac{1}{2} + iy$ on the line $\mathrm{Re}\ \tau = -\frac{1}{2}$ are clearly equivalent to the points $\frac{1}{2} + iy$ via the map $\tau \longmapsto \tau + 1$ induced by $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \mathrm{SL}_2(\mathbb{Z})$. Similarly, the points $-x + iy$, $0 \leq x \leq \frac{1}{2}$, on the piece of the unit circle bounding the fundamental domain are equivalent to the points $x + iy$ via the map $\tau \longmapsto -1/\tau$ induced by the matrix $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right) \in SL_2(\mathbb{Z})$.

By glueing these equivalent points together we get a topological space $Y(1)$ whose points are the orbits of $\mathbb{H}$ under the action of $\mathrm{SL}_2(\mathbb{Z})$. The space $Y(1)$ also inherits a complex analytic structure from $\mathbb{H}$, making it into a Riemann surface.

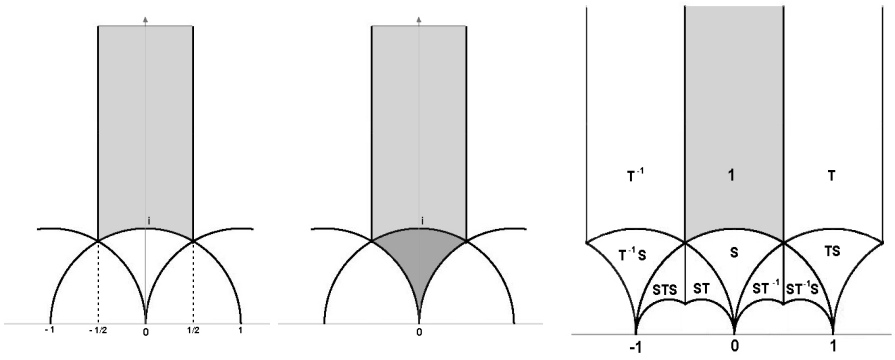

FIGURE 1. Fundamental domain for $\mathrm{SL}_2(\mathbb{Z})$

Note that instead of using the shaded part in the figure on the left as our fundamental domain, we could as well pick the dark shaded "triangle" in the middle figure. In fact, $\mathrm{SL}_2(\mathbb{Z})$ 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 the dark triangle is the result of applying the map $S : \tau \longmapsto -1/\tau$ to the original fundamental domain, which means that they are equivalent under the action of $\mathrm{SL}_2(\mathbb{Z})$. Observe that the point 0 on the boundary of the dark triangle corresponds to the point at infinity on the boundary of the standard fundamental domain. The figure on the right shows the action of some elements of $\mathrm{SL}_2(\mathbb{Z})$ on this fundamental domain.

This space $Y(1)$ can be made into a compact space $X(1)$ by adding a point (at infinity) and modifying the topology a little bit. We even can give $X(1)$ the structure of a complex analytic manifold, and it turns out that $X(1) \simeq \mathbb{P}^1\mathbb{C}$ is just the complex projective line, or the Riemann sphere.

2.3. **Level Structure.** It is an important observation that this classification of isomorphism classes of elliptic curves can be refined. In fact, consider a lattice Λ in $\mathbb{C}$ and its associated elliptic curve $E = \mathbb{C}/\Lambda$. Then $E[N] = \frac{1}{N}\Lambda/\Lambda$ is a group of type $\mathbb{Z}/N\mathbb{Z} \oplus \mathbb{Z}/N\mathbb{Z}$. We will use this group to 'add structure' to E.

- Consider pairs (E, C), where C is a cyclic subgroup of order N of $E[N]$; every point P of order N determines such a subgroup, but in general several P will generate the same C. We say that (E, C) and (E', C') are isomorphic if there is an isomorphism $\phi : E \longrightarrow E'$ of elliptic curves whose restriction to C satisfies $\phi(C) = C'$.
- Consider pairs (E, P), where P is a point of order N. Two pairs (E, P) and (E', P') are isomorphic if there exists an isomorphism $\phi : E \longrightarrow E'$ of elliptic curves with $\phi(P) = P'$.
- Consider triples (E, P, Q), where $\{P, Q\}$ is a basis for the group $E[N]$. We say that (E, P, Q) and (E', P', Q') are isomorphic if there exists an isomorphism $\phi : E \longrightarrow E'$ of elliptic curves with $\phi(P) = P'$ and $\phi(Q) = Q'$.

You may look at these variations as modifications of the category of elliptic curves whose morphisms are rational maps respecting the group law (we speak of adding a level-N structure): from the category of elliptic curves one builds e.g. the category whose objects are pairs (E, C), and whose morphisms $(E, C) \longrightarrow (E', C')$ are morphisms $f : E \longrightarrow E'$ in the category of elliptic curve with the additional property that the restriction of f to C maps C to C'.

Examples of such objects are easy to construct: to this end consider the elliptic curve $E_\tau = \mathbb{C}/\Lambda_\tau$ with $\Lambda_\tau = \mathbb{Z} \oplus \tau\mathbb{Z}$.

- The images of the points $\frac{1}{N}\mathbb{Z} + \tau\mathbb{Z}$ in E_τ form a subgroup C_τ of order N and gives a pair (E_τ, C_τ); the images of the points $\mathbb{Z} \oplus \frac{1}{N}\tau\mathbb{Z}$ define another group C' and another pair (E, C').
- The image P_τ of $\frac{1}{N}$ defines a pair (E, P_τ).
- The images P_τ and Q_τ of $\frac{1}{N}$ and $\frac{1}{N}\tau$ define a basis for $E_\tau[N]$ and give rise to an object (E_τ, P_τ, Q_τ).

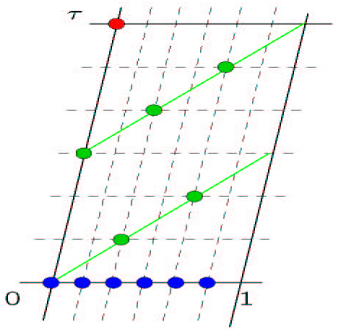

FIGURE 2. Level-6 structure on E_τ

This diagram[1] shows two level-6 structures on some elliptic curve E_τ, namely two different cyclic subgroups C of $E_\tau[6]$: the subgroup generated by the image of $\frac{1}{6}$ (blue points), and the subgroup generated by the image of $\frac{1}{3} + \frac{1}{2}\tau$ (green points).

[1]Stolen from Helena Verrill

3. Construction of the Riemann Surfaces

3.1. **Congruence Subgroups.** In order to describe isomorphism classes of these objects we need several subgroups of $\mathrm{SL}_2(\mathbb{Z})$, namely

- $\Gamma_0(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : c \equiv 0 \bmod N \right\}$;
- $\Gamma_1(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : a \equiv d \equiv 1,\ c \equiv 0 \bmod N \right\}$;
- $\Gamma(N) = \left\{ \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) : a \equiv d \equiv 1,\ b \equiv c \equiv 0 \bmod N \right\}$.

We clearly have $\Gamma(N) \subseteq \Gamma_1(N) \subseteq \Gamma_0(N) \subseteq \Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$. The canonical projection $\mathrm{SL}_2(\mathbb{Z}) \longrightarrow \mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z})$ has kernel $\Gamma(N)$, hence $\Gamma(N)$ is a normal subgroup of $\Gamma(1)$ with finite index $(\Gamma(1) : \Gamma(N)) = \#\,\mathrm{SL}_2(\mathbb{Z}/N\mathbb{Z}) = N^3 \prod_{p|N}(1 - p^{-2})$.

Now we can state

Theorem 2. *Let E be a complex elliptic curve with a cyclic subgroup C of order $N \geq 1$.*

- *There is a $\tau \in \mathbb{H}$ such that $(E, C) \simeq (E_\tau, C_\tau)$.*
- *We have $(E_\tau, C_\tau) \simeq (E_{\tau'}, C_{\tau'})$ for $\tau, \tau' \in \mathbb{H}$ if and only of there is some $M \in \Gamma_0(N)$ such that $M\tau = \tau'$.*

Thus isomorphism classes of pairs (E, C) correspond to points in $\Gamma_0(N)\backslash\mathbb{H}$.

Proof. Let E be a complex elliptic curve. Then $E \simeq \mathbb{C}/\Lambda$ for $\Lambda = \omega_1\mathbb{Z} \oplus \omega_2\mathbb{Z}$. We can choose the basis of the lattice in such a way that $C = \omega_1\mathbb{Z} \oplus \frac{1}{N}\omega_2\mathbb{Z}$. Put $\tau = \omega_1/\omega_2$ and define $u : \Lambda \longrightarrow \Lambda_\tau$ as multiplication by ω_2^{-1}. Then u induces an isomorphism $E \longrightarrow E_\tau$ that sends C to the group C_τ. □

Theorem 3. *Let E be a complex elliptic curve with a point P of order $N \geq 1$.*

- *There is a $\tau \in \mathbb{H}$ such that $(E, P) \simeq (E_\tau, P_\tau)$.*
- *We have $(E_\tau, P_\tau) \simeq (E_{\tau'}, P_{\tau'})$ for $\tau, \tau' \in \mathbb{H}$ if and only of there is some $M \in \Gamma_1(N)$ such that $M\tau = \tau'$.*

Thus isomorphism classes of pairs (E, P) correspond to points in $\Gamma_1(N)\backslash\mathbb{H}$.

Proof. Similar. □

Theorem 4. *Let E be a complex elliptic curve, and let $\{P, Q\}$ be a basis for $E[N]$ for some $N \geq 1$.*

- *There is a $\tau \in \mathbb{H}$ such that $(E, P, Q) \simeq (E_\tau, P_\tau, Q_\tau)$.*
- *We have $(E_\tau, P_\tau, Q_\tau) \simeq (E_{\tau'}, P_{\tau'}, Q_{\tau'})$ for $\tau, \tau' \in \mathbb{H}$ if and only of there is some $M \in \Gamma(N)$ such that $M\tau = \tau'$.*

Thus isomorphism classes of pairs (E, P, Q) correspond to points in $\Gamma(N)\backslash\mathbb{H}$.

Proof. Similar. □

Thus the isomorphism classes of our objects correspond to points on the associated Riemann surfaces $Y_0(N) = \Gamma_0(N)\backslash\mathbb{H}$, $Y_1(N) = \Gamma_1(N)\backslash\mathbb{H}$, and $Y(N) = \Gamma(N)\backslash\mathbb{H}$. The underlying topological spaces can be constructed by gluing together parts of the boundaries of the fundamental domains of these groups: as before, these are sets containing exactly one representative of $\mathbb{H}$ modulo the action of the congruence subgroups.

Computing these fundamental domains is rather straightforward: For some subgroup Γ with finite index in $\mathrm{SL}_2(Z)$, find a decomposition

$$\mathrm{SL}_2(\mathbb{Z}) = \bigcup_i \gamma_i \Gamma$$

of $\mathrm{SL}_2(\mathbb{Z})$ into cosets modulo Γ; if $\mathcal{F}$ denotes a fundamental domain of $\mathrm{SL}_2(\mathbb{Z})$, then

$$\mathcal{D} = \bigcup \gamma_i^{-1} \mathcal{F}$$

will be a fundamental domain for Γ.

For the level-2 subgroup $\Gamma = \Gamma_0(2)$, for example, it is easily shown that

$$\gamma_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad \gamma_3 = \begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$$

are representatives for $\mathrm{SL}_2(\mathbb{Z})/\Gamma_0(2)$; using the generators S and T of $\mathrm{SL}_2(\mathbb{Z})$, these elements can be written as $\gamma_1 = 1$, $\gamma_2 = S$, and $\gamma_3 = T^{-1}S^{-1}$. Now Figure 1 shows that the fundamental domain corresponding to this choice of representatives is given in Figure 3.

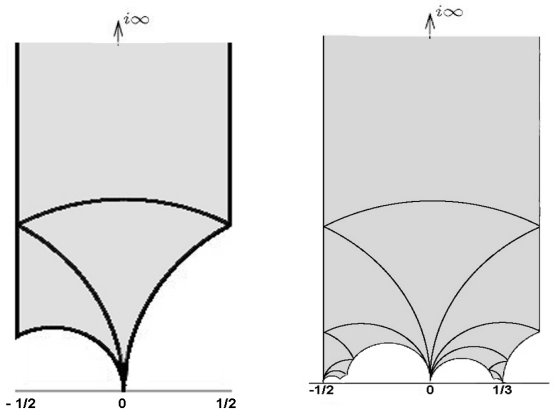

FIGURE 3. Fundamental domain for $\Gamma_0(2)$ and $\Gamma_0(6)$

As in the case of $Y(1)$, these Riemann surfaces are not compact but can be compactified in a natural way by adding several points called cusps; these compact Riemann surfaces $X_0(N)$, $X_1(N)$ and $X(N)$ then correspond, as every compact Riemann surface, to projective algebraic curves; these curves are called modular curves and play a central role in the theory of elliptic curves.

Note that the inclusions $\Gamma(N) \subseteq \Gamma_1(N) \subseteq \Gamma_0(N) \subseteq \Gamma(1)$ induce holomorphic maps (actually coverings of Riemann surfaces) $X(N) \longrightarrow X_1(N) \longrightarrow X_0(N) \longrightarrow X(1) = \mathbb{P}^1\mathbb{C}$. For example, the map $X_1(N) \longrightarrow X_0(N)$ sends the isomorphism class of (E, P) to the isomorphism class of (E, C), where $C = \langle P \rangle$ is the cyclic group of order N generated by P.

3.2. **Cusps.** "Compactification" sounds pretty simple, but recall that we not only have to give the spaces $X(N)$ a topology but also a complex structure: this actually requires some care.

The explanation of what the cusps are is not difficult, on the other hand: put $\mathbb{H}^* = \mathbb{H} \cup \mathbb{P}^1\mathbb{Q}$, i.e., add the rational points and infinity to the upper half plane. This can be given a natural topology, and then we put $X_0(N) = \Gamma_0(N)\backslash\mathbb{H}^*$, $X_1(N) = \Gamma_1(N)\backslash\mathbb{H}^*$, and $X(N) = \Gamma(N)\backslash\mathbb{H}^*$. Thus $X(N) \setminus Y(N)$ consists of the

equivalence classes of $\mathbb{P}^1\mathbb{Q}$ under the action of $\Gamma(N)$, and similarly for the other curves.

For the group $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$, there is only one equivalence class in $\mathbb{P}^1\mathbb{Q}$: given any fraction $\frac{d}{c}$ with $\gcd(c,d) = 1$, choose integers $a, b \in \mathbb{Z}$ with $ad - bc = 1$ (Bezout); then $M(\frac{d}{c}) = \infty$: thus all rational numbers are equivalent to the point at infinity on $\mathbb{P}^1\mathbb{Q}$, and adding this point to $Y(1)$ gives the Riemann sphere $X(1)$.

For primes p and $\Gamma_0(p)$, there are exactly $p+1$ cusps, and they can be computed rather easily from the decomposition of $\mathrm{SL}_2(\mathbb{Z})/\Gamma_0(N)$ into cosets.

Now that we have the compact Riemann surfaces, let me explain how to get from there to algebraic curves.

4. Modular Forms for $\mathrm{SL}_2(\mathbb{Z})$

4.1. **Modular Forms of Level N.** For $\tau \in \mathcal{H}$ and the associated lattice $\Lambda_\tau = \mathbb{Z} \oplus \tau\mathbb{Z}$, define the Eisenstein series

$$G_{2k}(\tau) = \sum_{(m,n)\neq(0,0)} \frac{1}{(m\tau + n)^{2k}}.$$

The Eisenstein series, viewed as functions on the upper half plane, have remarkable properties: let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ denote an element of $\mathrm{SL}_2(\mathbb{Z})$; then γ acts on the upper half plane via $\gamma z = \frac{az+b}{cz+d}$, and we have

$$G_{2k}(\gamma\tau) = (c\tau + d)^{2k} G_{2k}(\tau).$$

Meromorphic functions f on the upper half plane with the property

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

for all $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ are called **modular functions** of weight k; the Eisenstein series G_{2k} are modular functions of weight $2k$. They do not vanish at infinity: we have $G_{2k}(\infty) = 2\zeta(2k)$, where ζ denotes the Riemann zeta function.

Since $\mathrm{SL}_2(\mathbb{Z})$ is generated by the matrices $S = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$ and $T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, a function is modular of weight k if and only if

- $f(\tau + 1) = f(\tau)$ and
- $f(-1/\tau) = \tau^k f(\tau)$

for all $\tau \in bH$.

Let us introduce some notation: for a function $f : \mathbb{H} \longrightarrow \mathbb{C}$, some $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ and integers $k \in \mathbb{N}$ we define

$$(f|_k\gamma)(z) = (\det\gamma)^{-k/2}(cz + d)^{-k} f(\gamma z).$$

Then $(c_4|_4\gamma)(\tau) = c_4(\tau)$ and $(c_6|_6\gamma)(\tau) = c_4(\tau)$.

Meromorphic functions defined on $\mathbb{H}$ that are invariant under the translation $\tau \longmapsto \tau + 1$ (i.e. with respect to $\gamma = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$) have a Fourier expansion

$$f(\tau) = \sum_{n\in\mathbb{Z}} c_n q^n, \quad q = e^{2\pi i\tau}.$$

A modular function will be holomorphic at ∞ if $c_n = 0$ for all $n < 0$; such functions are called modular forms. Cusp forms are modular forms that vanish at infinity, i.e. that satisfy $c_0 = 0$. Modular forms and cusp forms of weight k clearly form vector spaces over $\mathbb{C}$ denoted by M_k and S_k. If k is odd, these spaces are empty. The modular forms of even weight are generated by the Eisenstein series: in fact, the products $G_4^m G_6^n$ are modular forms of weight $2k = 4m + 6n$, and every modular

form of weight $2k$ is a linear combination of these. In particular, $\dim M_k = 1$ for $k = 4, 6, 8, 10$ (here the M_k are generated by G_k), and $\dim M_{12} = 2$ since G_4^3 and G_6^2 are independent modular forms (in particular, G_{12} must be a linear combination of these two).

These results imply

$$\dim M_k = \begin{cases} \lfloor \frac{k}{12} \rfloor + 1 & \text{if } k \not\equiv 2 \bmod 12, \\ \lfloor \frac{k}{12} \rfloor & \text{if } k \equiv 2 \bmod 12. \end{cases}$$

Also $S_k = 0$ for $k < 12$; if $k \ge 12$ is even, then $\dim S_k = \dim M_k - 1$: this is because $M_k = S_k \oplus G_k\mathbb{C}$.

Eisenstein series can also be used to construct cusp forms: all we have to do is take two independent modular forms of weight $2k$ having the same value at infinity, and then take the difference. The smallest weight for which this works is $2k = 12$. Put $g_4 = 60G_4$ and $g_6 = 140G_6$; then $g_4(\infty) = \frac{4}{3}\pi^4$ and $g_6(\infty) = \frac{8}{27}\pi^6$, hence

$$\Delta = g_4^3 - 27g_6^2$$

is a modular form of weight 12 with $\Delta(\infty) = 0$, i.e., Δ is a cusp form of weight 12. Moreover, $\Delta(\tau) \ne 0$ for $\tau \in \mathbb{H}$.

Let us now construct some meromorphic functions on $X(1)$. The idea is simple: take any two modular forms $f_1, f_2 \in M_k$ of the same weight k; then f_1/f_2 is a meromorphic function of weight 0, hence invariant under the action of $\mathrm{SL}_2(\mathbb{Z})$ and therefore a meromorphic function not only on $\mathbb{H}$, but on $Y(1) = \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, and this extends to a meromorphic function on all of $X(1)$. Note that if $\dim M_k = 1$, then both functions will be constant multiples of each other, and we have constructed a constant function. Thus we only get nontrivial functions if we take k so large that $\dim M_k > 1$. This happens, as we have seen above, for $k = 12$, where we have the modular forms G_4^3 and G_6^2, as well as the cusp form Δ. This means that the j-function

$$j(\tau) = \frac{1728G_4^3}{\Delta}$$

is a modular function of weight 0, and since $\Delta(\tau) \ne 0$ for $\tau \in \mathbb{H}$, the only pole of j is at infinity.

The map $\tau \longmapsto [1728G_4^3 : \Delta]$ induces a map $j : X(1) \longrightarrow \mathbb{P}^1\mathbb{C}$; this map can be show to be bijective, hence the isomorphism classes of elliptic curves E_τ are in bijection with the values $j(\tau)$, or in other words: two elliptic curves E_τ, $E_{\tau'}$ are isomorphic if and only if $j(\tau) = j(\tau')$.

Let me note in passing that the q-expansions of Δ and j are given by

$$\begin{aligned} \Delta(\tau) &= q\prod_{n=1}^{\infty}(1-q^n)^{24} \\ &= q - 24q^2 + 252q^3 - 1472q^4 + \dots \\ j(\tau) &= \frac{1}{q} + 744 + 196884q + 21493760q^2 + \dots \end{aligned}$$

It is customary to write $\Delta(q) = \sum_{n=1}^{\infty}\tau(n)q^n$ and call $\tau(n)$ Ramanujan's tau function.

5. Modular Forms for Congruence Subgroups

More generally, we can consider modular forms of level N, weight k, and nebentypus χ: these are functions on $\mathbb{H}$ such that

(1) f is holomorphic in $\mathbb{H}$;
(2) $(f|_k\gamma)(z) = \chi(d)f(z)$ for all $z \in \mathbb{H}$ and $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$; here χ is a Dirichlet character whose conductor f divides the level N;
(3) f is holomorphic at the cusps.

The set of all modular forms of level N, weight k, and nebentypus χ forms a $\mathbb{C}$-vector space and is denotes by $M_k(N, \chi)$. If χ is the trivial character, we write $M_k(N)$.

A modular form f is called a cusp form if f vanishes at the cusps. The $\mathbb{C}$-vector space of cusp forms of level N, weight k, and nebentypus χ is denoted by $S_k(N, \chi)$.

Proposition 5. *We have* $\dim S_2(N) = g(X_0(N))$*, the genus of* $X_0(N)$*.*

This is because the map $f(z) \longmapsto 2\pi i f(z)dz$ induces a $\mathbb{C}$-linear map between the space of cusp forms of weight 2 and the space $\Omega^1(X_0(N))$ of holomorphic differential forms on $X_0(N)$. The reason for this is that $\frac{d\gamma(z)}{dz} = (cz + d)^{-2}$ by simple calculus, hence the differential forms are invariant under the action of γ, (i.e. $f(\gamma(z))d\gamma(z) = f(z)dz$) if and only if f has weight 2.

Here comes the deep stuff:

Theorem 6. *The curves* $X_1(N)$ *and* $X_0(N)$ *are defined over* $\mathbb{Q}$*. The spaces of modular and cusp forms* $M_k(N)$ *and* $S_k(N)$ *have bases with integral Fourier coefficients.*

Examples:

$$X_0(11) : y^2 + y = x^3 - x^2 - 10x - 20;$$
$$X_1(11) : y^2 + y = x^3 - x^2$$

Once we know that the curves $X_1(N)$ and $X_0(N)$ have equations defined over $\mathbb{Q}$, we can ask about their reductions modulo primes p. A deep theorem in this direction is

Theorem 7. *The reductions modulo* p *of* $X_1(N)$ *and* $X_0(N)$ *are nonsingular for all primes* $p \nmid N$*.*

5.1. **Sums of four squares.** Modular forms also occur in other parts of number theory. Consider the function

$$\theta(\tau) = \sum_{n \in \mathbb{Z}} q^{n^2}, \quad \text{where } q = e^{2\pi i \tau}.$$

Then

$$\theta(\tau)^4 = \sum_{n=0}^{\infty} r_4(n) q^n,$$

where $r_4(n)$ is the number of $(x_1, x_2, x_3, x_4) \in \mathbb{Z}^4$ with $n = x_1^4 + x_2^4 + x_3^4 + x_4^4$. The proof of Jacobi's theorem that

$$r_4(n) = \begin{cases} 8 \sum_{d|n} d & \text{if } n \equiv 1 \bmod 2, \\ 24 \sum_{d|n,\ 2 \nmid d} d & \text{if } n \equiv 0 \bmod 2, \end{cases}$$

is based on this identity and on analytic properties of θ, namely the fact that it is a modular form of weight 2 with respect to $\Gamma_0(4)$.

6. The Modular Equation

The curves $X_1(N)$ and $X_0(N)$ are defined over $\mathbb{Q}$, and so is the natural projection $X_1(N) \longrightarrow X_0(N)$; this means that rational points on $X_1(N)$ will get mapped to rational points on $X_0(N)$. For some levels, this can be used to show that $X_1(N)$ does not have any rational points by showing that this holds for the curve $X_0(N)$, which has smaller genus and therefore is often easier to deal with.

Now it is very easy in theory to come up with an explicit equation for $X_0(N)$: it can be shown that the function $j_N(\tau) = j(N\tau)$ is algebraic over $\mathbb{C}(j)$, hence there is a polynomial $\Phi_N(x,y)$ with $\Phi_N(j,j_N) = 0$. This polynomial has coefficients in $\mathbb{Z}$, and it defines the affine part of $X_0(N)$. Unfortunately its coefficients as well as its degree are so large (in particular, the equation is highly singular, so in order to get better equations you have to remove singularities) that this is almost impossible even for $N = 11$.

A practical method for computing models of $X_0(N)$ is the following: compute a basis $\{f_1, \dots, f_g\}$ of the space $S_2(N)$ of cusp forms of weight 2; we know that we can choose a basis whose q-coefficients are integers. This space has dimension g, where g is the genus of $X_0(N)$. Now

$$X_0(N) \longrightarrow \mathbb{P}^{g-1}(\mathbb{C}) : \tau \longmapsto [f_1(\tau) : \cdots f_g(\tau)]$$

is a well defined map away from the cusps: this is because all the f_i behave in the same way (that is, get multiplied by the same factor) when acted on by some $\gamma \in \Gamma_0(N)$. If $g \ge 2$ and $X_0(N)$ is not hyperelliptic, then this map is injective, and now relations between the f_i then give equations describing the image of $X_0(N)$ in the projective space of dimension $g-1$. If $g = 1$ or if $X_0(N)$ is hyperelliptic, there are other methods for getting an equation.

Nevertheless, here are a few q-expansions for cusp forms of weight 2 for $\Gamma_0(N)$ for the smallest values of N with genus 1:

N	f
11	$q - 2q^2 - q^3 + 2q^4 + q^5 + 2q^6 + \dots$
14	$q - q^2 - 2q^2 + q^3 + 2q^6 + \dots$
15	$q - q^2 - q^3 - q^4 + q^5 + q^6 + \dots$
17	$q - q^2 - q^4 - 2q^5 + \dots$

The relations between the f_i give equations in $\mathbb{P}^{g-1}\mathbb{Q}$ describing the curve; by eliminating variables using resultants it is possible to realize these curves as plane affine curves, but doing so often produces singularities. This cannot be avoided, of course: the genus formula shows that there cannot be a smooth plane algebraic curve with genus 2, for example.

Some meromorphic functions on $X_0(N)$ can be constructed directly. In fact, consider Dedekind's eta function

$$\eta(\tau) = q^{1/24} \prod_{n=1}^{\infty} (1 - q^n),$$

where $q = e^{2\pi i\tau}$ and $q^{1/24} = e^{2\pi i\tau/24}$, and put $\eta_k(\tau) = \eta(k\tau)$.

Sometimes products of powers of these eta functions, such as $h = \eta^2\eta_{11}^2$, turn out to be nice functions:

$$h(\tau) = \eta(\tau)^2\eta_{11}(\tau)^2 = q\prod_{n=1}^{\infty}(1-q^n)^2(1-q^{11n})^2 = q - 2q^2 - q^3 + 2q^4 + \dots$$

has a q-expansion and is therefore invariant under $\tau \longmapsto \tau + 1$. It turns out that h actually is a cusp form of weight 2 for $\Gamma_0(11)$. Using eta functions, it is possible to write down modular forms for quite a few levels; for example, $\eta^2\eta_2^2\eta_3^2\eta_6^2 = q - 2q^2 - 3q^3 + 4q^4 + 6q^5 + \dots$ is a cusp form of weight 4 for $\Gamma_0(6)$. Note, however, that not every modular form can be constructed in this way.

We get more freedom by allowing negative exponents in such products, but then the result will not be holomorphic in general. But in some cases such a product of eta functions is modular of weight 0, and the following theorem due to Ligozat tells us exactly when this happens:

Theorem 8. *The function $\prod_{d|N}\eta_d^{r_d}$ is a modular function of weight 0 for $\Gamma_0(N)$ (in other words: an element of the function field of $X_0(N)$) if and only if the following conditions are satisfied:*

1. $\sum_{d|N} r_d = 0$;
2. $\sum_{d|N} r_d d \equiv 0 \bmod 24$;
3. $\sum_{d|N} r_d \frac{N}{d} \equiv 0 \bmod 24$;
4. $\prod_{d|N}(N/d)^{r_d}$ *is a square in* $\mathbb{Q}$.

For example, $h_6 = \eta_1^5\eta_3\eta_2^{-1}\eta_6^{-5}$ is a modular function of weight 0 with respect to $X_0(6)$, and its q-expansion is $h_6 = \frac{1}{q} - 5 + 10q - 16q^2 + \dots$.

Here are some hyperelliptic modular curves:

N	$X_1(N)$
13	$y^2 = x^6 + 4x^5 + 6x^4 + 2x^3 + x^2 + 2x + 1$
16	$y^2 = x^6 + 2x^5 - x^4 - x^2 - 2x + 1$
18	$y^2 = x^6 + 2x^5 + 5x^4 + 10x^3 + 10x^2 + 4x + 1$
28	$y^2 = x^5 - 4x^4 - 13x^3 - 9x^2 - x$
40	$y^2 = x^5 - 3x^4 - 12x^2 - 16x$

7. Mazur's Theorem

Theorem 9. *If the curve $X_0(N)$ has genus $g > 0$, then its only rational points are the rational cusps.*

The modular curves $X_1(N)$ with genus > 0 are those with $N = 11$ and $N \geq 13$.

Not all cusps are rational: the curve $X_1(11)$ has five rational cusps and five cusps defined over the maximal real subfield of $\mathbb{Q}(\zeta_{11})$.

Recall that the points of $X_1(N)$ (with the exception of cusps) correspond to isomorphism classes (E, P), where P is a point of order N. It turns out that the isomorphism classes of pairs (E, P), where E is an elliptic curve defined over $\mathbb{Q}$ and $P \in E(\mathbb{Q})$ a rational point of order N, correspond exactly to the non-cuspidal rational points on $X_1(N)$. Thus in order to prove that there are no elliptic curves defined over $\mathbb{Q}$ with a rational torsion point of order N, all one has to do is compute an explicit equation for $X_1(N)$ defined over $\mathbb{Q}$ and show that its only rational points are cusps. Mazur's proof, of course, does not use explicit equations of modular curves.

Exactly the same thing works for $X_0(N)$: its non-cuspidal rational points correspond to pairs (E,C) of elliptic curves defined over $\mathbb{Q}$ and rational cyclic subgroups C of order N; here a subgroup C is called rational if it is a $\mathrm{Gal}\,(\overline{\mathbb{Q}}/\mathbb{Q})$-module, in other words: if every automorphism of $\overline{\mathbb{Q}}/\mathbb{Q}$ maps elements of C to other elements of C. This does not mean that all the points in C are rational (this is a sufficient condition for C to be rational, but not necessary): consider the example $E: y^2 = x^3 - 1$; letting ρ denote a primitive cube root of unity, the subgroup $C = \{\mathcal{O}, (\rho,1), (\rho^2,1)\}$ of $E[3]$ is rational!

Mazur also classified all the possible N for which there exist elliptic curves defined over $\mathbb{Q}$ with a rational cyclic subgroup of order N; the possible values of N are $1 \le N \le 10$, $N = 12$ (of course), as well as $N = 13, 16, 18, 25$.

7.1. $X(2)$**.** As an example, consider the group $\Gamma(2)$ of matrices in $\mathrm{SL}_2(\mathbb{Z})$ with a, d odd and b, c even. The curve $X(2)$ is a curve of genus 0, i.e., the projective line. It has three cusps represented by the elements $0, 1, \infty \in \mathbb{P}^1\mathbb{Q}$.

The non-cuspidal points $\lambda \in \mathbb{P}^1\mathbb{C}$ parametrize elliptic curves with a 2-torsion group $E[2] \simeq \mathbb{Z}/2 \oplus \mathbb{Z}/2$. Its rational non-cuspidal points $\lambda \in \mathbb{P}^1\mathbb{Q}$ parametrize elliptic curves defined over $\mathbb{Q}$ with three rational points of order 2. In fact, such a parametrization is given by the Legendre family $E_\lambda: y^2 = x(x-1)(x-\lambda)$.

Note that the cusps $\lambda = 0$ and $\lambda = 1$ do not correspond to elliptic curves, but to singular cubics.