Author: Derickx, M.
Title: Torsion points on elliptic curves over number fields of small degree
Issue Date: 2016-09-21

CHAPTER 1

Modular curves and modular forms

MODULAR CURVES AND MODULAR FORMS

MAARTEN DERICKX

Contents

1. Elliptic curves

There are two ways in which one can look at modular curves, one is the standpoint of complex geometry and the other is the standpoint of algebraic geometry. These two standpoints meet at the place where one starts to do algebraic geometry over $\mathbb{C}$. In this section the theory of both sides is discussed in parallel.

A complex elliptic curve is pair $(E, 0)$ of a compact Riemann surface E of genus 1 together with a base point $0 \in \mathbb{C}$. In both the complex and the algebraic setting we will just write E instead of $(E, 0)$ in the rest of this text.

Let $\Lambda \subset \mathbb{C}$ be a lattice, i.e. a discrete subgroup of maximal rank, meaning rank 2 in this case. Then

$$E_\Lambda := \mathbb{C}/\Lambda$$

together with the equivalence class of $0 \in \mathbb{C}$ is an elliptic curve. The holomorphic one form $\mathrm{d}z$ on $\mathbb{C}$ is invariant under translation by Λ and hence descends to a nonzero holomorphic one form on $\mathbb{C}/\Lambda$.

Conversely if $\omega \in \Omega^1(E)$ is a nonzero holomorphic one form, then there exists a unique lattice $\Lambda_{E,\omega} \subset \mathbb{C}$ and a unique isomorphism

$$f : E \xrightarrow{\sim} \mathbb{C}/\Lambda_{E,\omega}$$

such that $f^*(\mathrm{d}z) = \omega$.

Using the isomorphism f, the elliptic curve E gets a group law, the group law is independent of the choice of $\mathrm{d}z$ since scalar multiplication $\mathbb{C} \to \mathbb{C}$ is a group homomorphism.

Let S be a scheme, an algebraic elliptic curve over S is a pair $(E, 0)$ where E is a scheme that is smooth of relative dimension 1 and proper over S such that all its geometric fibers are irreducible genus one curves and $0 \in E(S)$.

Let R be a commutative $\mathbb{Z}[\frac{1}{6}]$-algebra, and $a_4, a_6 \in R$ such that

$$-16(4a_4^3 + 27a_6^2) \in R^*,$$

then the projective curve E_{a_4,a_6} given by

$$y^2 = x^3 + a_4x + a_6$$

together with ∞ is an elliptic curve and

$$\omega_{a_4,a_6} := (3x^2 + a_4)^{-1}\mathrm{d}y = (2y)^{-1}\mathrm{d}x$$

is a global one form.

Suppose $\operatorname{Spec} R \subset S$ is an affine open with $6 \in R^*$ such that there exist a nowhere vanishing 1 form $\omega \in \Omega^1_{E/R}(E)$ then there are unique $a_4, a_6 \in R$ and a unique

$$f : E \xrightarrow{\sim} E_{a_4,a_6}$$

such that $f^*\omega_{a_4,a_6} = \omega$

One can put a group scheme structure on E by dentifying E with $\mathrm{Pic}^0_{E/S}$ by sending $P \in E(T)$ to the line bundle T $\mathcal{O}_{E_T}(P - 0_T)$ for all S schemes T.

The story on the algebraic side can be generalized to the case of $\mathbb{Z}$-algebras with a little bit more effort. Also one can mimic the definition of the group structure in the complex case by first putting a group scheme structure on E_{a_4,a_6} using explicit equations, and use f to give E a group scheme structure as well. So we have seen that both complex and algebraic elliptic curves, although defined by abstract properties can always be written down explicitly, and that they automatically inherit a group (scheme) structure.

If E_a is an algebraic elliptic curve over $\mathbb{C}$ then $E_a(\mathbb{C})$ is a complex elliptic curve and if E_c is a complex elliptic curve then one write $E_c \cong \mathbb{C}/\Lambda$. Define

$$\begin{aligned} \wp_\Lambda : \mathbb{C} \setminus \Lambda &\to \mathbb{C} && (1)\\ z &\mapsto \frac{1}{z^2} + \sum_{\lambda \in \Lambda \setminus \{0\}} \left(\frac{1}{(z-\lambda)^2} - \frac{1}{(-\lambda)^2} \right) \\ G_{2k}(\Lambda) &:= \sum_{\lambda \in \Lambda \setminus \{0\}} \frac{1}{\lambda^{2k}}, \quad \text{for } k \in \mathbb{Z}_{\geq 2} && (2)\\ g_2(\Lambda) &:= 60 G_4(\Lambda), \quad g_3(\Lambda) := 140 G_6(\Lambda). && (3) \end{aligned}$$

The function $\wp_\Lambda$ is called the Weierstrass P-function. The function $\wp_\Lambda$ and its derivative satisfy the following equation

$$(\tfrac{1}{2}\wp'_\Lambda(z))^2 = \wp_\Lambda(z)^3 - \tfrac{1}{4} g_2(\Lambda)\wp_\Lambda(z) - \tfrac{1}{4} g_3(\Lambda).$$

The functions $\wp_\Lambda$ and $\wp'_\Lambda$ are invariant under translation by Λ so they induce a map

$$\begin{aligned} f_\Lambda : \mathbb{C}/\Lambda &\to E_{\frac{1}{4}g_2, \frac{1}{4}g_3}(\mathbb{C}) && (4)\\ z &\mapsto \wp_\Lambda(z), \tfrac{1}{2}\wp'_\Lambda(z), \end{aligned}$$

where the equivalence class $0 + \Lambda$ is sent to ∞. The map f_Λ is an isomorphism of elliptic curves, and it is even compatible with the choice of one forms since

$$f_\Lambda^*(\frac{\mathrm{d}x}{2y}) = \frac{\mathrm{d}\wp(z)}{\wp'(z)} = \mathrm{d}z. \qquad (5)$$

Two elliptic curves $\mathbb{C}/\Lambda_1$ and $\mathbb{C}/\Lambda_2$ are isomorphic if an only if there exists a u in $\mathbb{C}^*$ such that $\Lambda_2 = u\Lambda_1$. Define the j-invariant of $\mathbb{C}/\Lambda$ by

$$j(\Lambda) := 1728 \frac{g_2(\Lambda)^3}{g_2(\Lambda)^3 - 27 g_3(\Lambda)^2}.$$

Using the fact that $g_2(u\Lambda) = u^{-4} g_2(\Lambda)$ and $g_3(u\Lambda) = u^{-6} g_3(\Lambda)$ it follows that j only depends on the isomorphism class of $\mathbb{C}/\Lambda$ and one can even show that j determines the isomorphism class uniquely.

Let R be a $\mathbb{Z}[\frac{1}{6}]$ algebra and $a_4, a_6, a'_4, a'_6 \in R$ such that E_{a_4,a_6} and $E_{a'_4,a'_6}$ are elliptic curves. These curves are isomorphic over R if and only if there exists an $u \in R^*$ such that $a'_4 = u^{-4} a_4$ and $a'_6 = u^{-6} a_6$. Define

$$j(a_4, a_6) := 1728 \frac{4a_4^3}{4a_4^3 + 27 a_6^2}.$$

Then $j(a_4, a_6)$ only depends on the isomorphism class of E_{a_4,a_6} and if R is an algebraically closed field then j even determines it uniquely.

This shows that both the complex and the algebraic way of looking at elliptic curves agree, if in the algebraic world one restricts to elliptic curves over $\mathbb{C}$.

1.1. **Some q-expansions.** Two elliptic curves $\mathbb{C}/\Lambda$ and $\mathbb{C}/\Lambda'$ are isomorphic if and only if there exists a $c \in \mathbb{C}$ such that $c\Lambda = \Lambda'$. Now chose two generators λ_1, λ_2 of Λ. By scaling with λ_2^{-1} one sees that $\mathbb{C}/\Lambda$ is isomorphic to $\mathbb{C}/(\lambda_1/\lambda_2\mathbb{Z} + \mathbb{Z})$. In particular, by replacing λ_1/λ_2 by $-\lambda_1/\lambda_2$ if necessary, one sees that there is always a $\tau \in \mathbb{H} := \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\}$ such that $\mathbb{C}/\Lambda \cong \mathbb{C}/\tau\mathbb{Z} + \mathbb{Z}$. For $\tau \in \mathbb{H}$ define $\Lambda_\tau := \tau\mathbb{Z} + \mathbb{Z}$. By additionally defining $\wp(z, \tau) := \wp_{\Lambda_\tau}(z)$, $G_{2k}(\tau) := G_{2k}(\Lambda_\tau)$ and $g_i(\tau) = g_i(\Lambda)$ for $i = 2, 3$ one can view $\wp$ as a meromorphic function on $\mathbb{C} \times \mathbb{H}$ and G_{2k} and g_i as holomorphic functions on $\mathbb{H}$. All these functions are invariant under translation by 1 on the τ coordinate since Λ_τ and $\Lambda_{\tau+1}$ are the same lattice. Also z and $z + 1$ define the same point in $\mathbb{C}/\Lambda_\tau$ showing that $\wp$ is also invariant under translation by 1 in the z coordinate. This means that all these functions can be written as power series in $q := e^{2\pi i\tau}$ whose coefficients are Laurent series in $u := e^{2\pi iz}$. See for example [Silverman(1994), I §6,§7]. The resulting power series are

$$\wp(z, \tau) = (2\pi i)^2 \left(\sum_{n\in\mathbb{Z}} \frac{q^n u}{(1 - q^n u)^2} + \frac{1}{12} - 2\sum_{n=1}^{\infty} \frac{q^n}{(1 - q^n)^2} \right) \tag{6}$$

$$G_{2k}(\tau) = (2\pi i)^{2k} \left(\frac{-B_{2k}}{(2k)!} + \frac{2}{(2k-1)!} \sum_{n=1}^{\infty} \frac{n^{2k-1} q^n}{1 - q^n} \right), \tag{7}$$

where $B_k \in \mathbb{Q}$ are the Bernoulli numbers, which are defined as the coefficients of the Taylor series $\frac{t}{e^t - 1} = \sum_{k=1}^{\infty} B_k \frac{t^k}{k!}$. Applying $\frac{\partial}{\partial z} = 2\pi i u \frac{\partial}{\partial u}$ to $\wp$ one obtains the formula[1]

$$\frac{\partial \wp(z, \tau)}{\partial z} := -(2\pi i)^3 \sum_{n\in\mathbb{Z}} \frac{q^n u(1 + q^n u)}{(1 - q^n u)^3} \tag{8}$$

The formula's for $G_{2k}(\tau)$ and $g_i(\tau)$ are often rewritten using the auxiliary functions

$$\sigma_k(n) := \sum_{d|n, d>0} d^k, \quad s_k(q) := \sum_{n=1}^{\infty} \frac{n^k q^n}{1 - q^n} = \sum_{n=1}^{\infty} \sigma_k(n) q^n. \tag{9}$$

One has $B_4 = -\frac{1}{30}$ and $B_6 = \frac{1}{42}$ so that the q-expansion of $\frac{1}{4}g_2(\tau)$ and $\frac{1}{4}g_3(\tau)$ are

$$\frac{1}{4}g_2(\tau) := (2\pi i)^4 (\frac{1}{48} + 5s_3(q)) \text{ and } \frac{1}{4}g_3(\tau) := (2\pi i)^6 (-\frac{1}{864} + \frac{7}{12}s_5(q)). \tag{10}$$

1.2. **Tate Curve.** Let τ be in the upper half plane, then the elliptic curve $y^2 = x^3 - \frac{1}{4}g_2(\tau)x - \frac{1}{4}g_3(\tau)$ has j-invariant $j(\tau) := j(\Lambda_\tau)$ and discriminant $\Delta(\tau) : g_2(\tau)^3 - 27g_3(\tau)^2$. Using the above formulas for q-expansion one can show that

$$j(\tau) = \frac{1}{q} + \sum_{n=0}^{\infty} c(n) q^n, \quad c(n) \in \mathbb{Z}, \text{ and} \tag{11}$$

$$\Delta(\tau) = (2\pi i)^{12} q \prod_{n=1}^{\infty} (1 - q^n)^{24} \tag{12}$$

[1] This differs by a minus sign from the formula in [Silverman(1994), I Thm 6.2], where there is a sign mistake.

Define $\tilde{g}_2 = (2\pi i)^{-4} g_2$, $\tilde{g}_3 := (2\pi i)^{-6} g_3$, $\tilde{\wp} := (2\pi i)^{-2}\wp$, and $\tilde{\Delta} := (2\pi i)^{-12}\Delta$. With these definitions the elliptic curve $y^2 = x^3 - \frac{1}{4}\tilde{g}_2(\tau)x - \frac{1}{4}\tilde{g}_3(\tau)$ is isomorphic to $\mathbb{C}/\Lambda_\tau$ via $(x, y) = (\tilde{\wp}(z, \tau), \frac{1}{2}\frac{\partial}{\partial z}\tilde{\wp}(z, \tau))$. This model of $\mathbb{C}/\Lambda_\tau$ over $\mathbb{H}$ has not only a j-invariant whose q-expansion has integral coefficients, but the coefficients of the q-expansion of its discriminant $\tilde{\Delta}$ are integral as well. The functions $\tilde{g}_2$ and $\tilde{g}_3$ do not have integral q-expansions, although they are almost integral since they are fractions whose denominator is a divisor of $864 = 2^3 \cdot 3^3$ as formula (10) shows. Substituting $x = x' + \frac{1}{12}$ and $y = y' + \frac{1}{2}x'$ gives the curve

$$y'^2 + x'y' = x'^3 + a_4x' + a_6, \quad a_4 := -5s_3, \quad a_6 := -\frac{5s_3 + 7s_5}{12} \tag{13}$$

It is clear that a_4 has integral coefficients in its q-expansion. For any integer n one has $5n^3 + 7n^5 \equiv 0 \mod 12$ so that a_6 also has integral coefficients. The *Tate curve* E_q is the curve (13) over $\mathbb{Z}[[q]]$ where one uses q-expansion to see a_4 and a_6 as elements of $\mathbb{Z}[[q]]$. It is not an elliptic curve since its fiber above $q = 0$ is singular, however since $\tilde{\Delta}$ is a unit in $\mathbb{Z}[[q]][\frac{1}{q}]$ it is an elliptic curve over $\mathbb{Z}[[q]][\frac{1}{q}]$. The Tate curve is useful since it allows one to study elliptic curves over p-adic fields, i.e. finite extensions of $\mathbb{Q}_p$. This is captured in the following Theorem due to Tate whose statement can be obtained by combining [Silverman(1994), V Thm 3.1 and Lemma 5.1]

Theorem 1.1 (Tate). *Let K be a p-adic field and $q_0 \in K^*$ with $|q_0| < 1$ then the power series a_4 and a_6 converge in q_0. Let E_{q_0} be the curve given by*

$$y'^2 + x'y' = x'^3 + a_4(q_0)x' + a_6(q_0)$$

then $E_{q_0}(\overline{K})$ is isomorphic to $\overline{K}^*/q_0$ as $\mathrm{Gal}(\overline{K}/K)$ modules. The curve E_{q_0} has $|j(E_{q_0})| > 1$ and for every elliptic curve E over K with $|j(E_{q_0})| > 1$ there is a unique $q_0 \in K$ such that $E \cong E_{q_0}$ over $\overline{K}$.

The isomorphism between $\overline{K}^*/q_0$ and $E_{q_0}(\overline{K})$ is obtained by using formula's 6 and 8 to find the q-expansions of $x' = \tilde{\wp} - \frac{1}{12}$ and $y' = \frac{1}{2}\frac{\partial}{\partial z}\tilde{\wp} - \frac{1}{2}\tilde{\wp} - \frac{1}{12}$. With this isomorphism one sees that the invariant differentials

$$2\pi i \mathrm{d}z = \frac{\mathrm{du}}{u} = \frac{\mathrm{d}x'}{2y' + x'},$$

are equal, where the left most differential only makes sense in the complex world. The above theorem is the p-adic analogue of the fact that every elliptic curve over $\mathbb{C}$ can be written as $\mathbb{C}/(\tau\mathbb{Z} + \mathbb{Z}) \cong \mathbb{C}^*/e^{2\pi i\tau^{\mathbb{Z}}}$.

1.3. **Néron polygons.** The fiber of the Tate curve E_q over $\mathbb{Z}[[q]]$ at $q = 0$ is not an elliptic curve although it is still a curve, in fact it's special fiber is isomorphic to $\mathbb{P}^1$ with two points glued together. The special fiber is an example of a Néron 1-gon. In general if N is an integer and R is a ring then the *Néron N-gon $\mathcal{N}_N$* over R is defined to be the singular projective curve over R that one obtains by taking a copy X_i of $\mathbb{P}^1_R$ for each $i \in \mathbb{Z}/N\mathbb{Z}$ and glueing the point ∞ of X_i to the point 0 of X_{i+1} in such a way that the intersections become ordinary double points. Using the identification $\mathbb{G}_{m,R} = \mathbb{P}^1_R \setminus \{0, \infty\}$ one sees that the smooth locus of the Néron N-gon over R is isomorphic to $\mathbb{Z}/N\mathbb{Z} \times \mathbb{G}_{m,R}$, turning the smooth locus of the Néron N-gon into a group scheme. Morphisms between Néron N-gons are the scheme morphisms that induce group-scheme homomorphisms when restricted to the smooth locus, so in particular they should map the smooth locus to itself. If K is a field of characteristic co-prime to N then one can make $\mu_N(K)$ act on $\mathbb{Z}/N\mathbb{Z} \times \mathbb{P}^1_K$ by $\zeta_N(i,(a:b)) := (i,(\zeta_N^i a : b))$ and one can make $\{\pm 1\}$ act on it by $-(i,(a:b)) := (i,(b:a))$. Both these actions are group homomorphisms when restricted to $\mathbb{Z}/N\mathbb{Z} \times \mathbb{G}_{m,K} \subseteq \mathbb{Z}/N\mathbb{Z} \times \mathbb{P}^1_K$ and they are compatible with the identifications of ∞ on the i-th component with 0 on the $i+1$-th component. Since these actions commute, one gets that automorphism group of $\mathcal{N}_N$ contains

$$\mu_N(K) \times \{\pm 1\}.$$

The above group is actually the entire automorphism group.

1.4. **Generalized Elliptic curves.** Theorem 1.1 shows that the Tate curve E_q over $\mathbb{Z}[[q]]$ can be used to study elliptic curves over p-adic fields with $|j| > 1$ and $j \neq \infty$. Its special fiber at $q = 0$ is not an elliptic curve but it is still a Néron N-gon. Generalized elliptic curves are curves where we also allow the geometric fibers to be Néron N-gons, to be more precise.

Definition 1.2. Let S be a scheme, a *generalized elliptic curve* over S is a scheme E that is proper, flat and of finite presentation over S together with a group scheme structure on E^{sm}, such that each of the geometric fibers $E_{\overline{K}}$ of E is isomorphic to either an elliptic curve over K or the Néron N-gon over K.

In the above definition E^{sm} denotes the locus of E that is smooth over S and the isomorphisms of the geometric fibers should respect the group scheme structure on $E^{sm}_{\overline{K}}$. A point of order N on a generalized elliptic curve E/S is understood to be an element $P \in E(S)$ of order N such that all geometric fibers of P also have order N and furthermore such that the subgroup generated by P meets all components of all geometric fibers.

2. Modular curves

2.1. **The modular curve $Y_1(N)$.** Modular curves are curves whose points correspond to elliptic curves with some extra structure. The modular curve $Y_1(N)$ is the curve whose points correspond to an elliptic curve with a torsion point of order N. To avoid technical difficulties we assume that $N > 4$ is an integer. Let (E_1, P_1) and (E_2, P_2) be pairs of an elliptic curve together with a point of order N, then an isomorphism from (E_1, P_1) to (E_2, P_2) is defined to be an isomorphism of elliptic curves $f : E_1 \to E_2$ such that $f(P_1) = f(P_2)$. This definition will be used for both complex and algebraic elliptic curves.

Let $\mathbb{H}$ be the complex upper half plane. To $\tau \in \mathbb{H}$ one can associate the elliptic curve $E_\tau := \mathbb{C}/(\tau\mathbb{Z} + \mathbb{Z})$. If $E = \mathbb{C}/\Lambda$ is an elliptic curve and ω_1, ω_2 are generators of Λ such that $\operatorname{Im}(\omega_1/\omega_2) > 0$ then division by ω_2 gives an isomorphism $E \cong E_{\omega_1/\omega_2}$ showing that every elliptic curve is isomorphic to some E_τ.

Let $\mathrm{SL}_2(\mathbb{Z})$ act on $\mathbb{H}$ by $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]\tau = \frac{a\tau+b}{c\tau+d}$. Then the sequence of isomorphisms

$$\begin{aligned} E_\tau &\cong \mathbb{C}/\left((a\tau+b)\mathbb{Z} + (c\tau+d)\mathbb{Z}\right) \\ &\cong \mathbb{C}/\left(\tfrac{a\tau+b}{c\tau+d}\mathbb{Z} + \mathbb{Z}\right) = E_{\frac{a\tau+b}{c\tau+d}}. \end{aligned}$$

shows that if $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, then $E_{\gamma\tau} \cong E_\tau$. One can even show that if $\tau_1, \tau_2 \in \mathbb{H}$ then $E_{\tau_1} \cong E_{\tau_2}$ if and only if there exists a $\gamma \in \mathrm{SL}_2(\mathbb{Z})$ such that $\tau_2 = \gamma\tau_1$.

The point $\frac{1}{N} \in E_\tau$ has order N, and because $N > 3$ one can show that E_τ has no automorphisms that fix $\frac{1}{N}$. Now

$$\tfrac{c\tau+d}{N} \equiv \tfrac{1}{N} \mod \tau Z + \mathbb{Z}$$

if and only if $(c,d) \equiv (0,1) \mod N$, so if one defines $\Gamma_1(N) \subseteq \mathrm{SL}_2(\mathbb{Z})$ to be the set of matrices with $(c,d) \equiv (0,1) \mod N$ then the isomorphism

$$\left(E_{\frac{a\tau+b}{c\tau+d}}, \tfrac{1}{N}\right) \cong \left(E_\tau, \tfrac{c\tau+d}{N}\right)$$

shows that if $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, then $(E_{\gamma\tau}, 1/N) \cong (E_\tau, 1/N)$ if and only if $\gamma \in \Gamma_1(N)$. So that $\mathbf{Y}_1(N) := \Gamma_1(N)\backslash\mathbb{H}$ can be interpreted as the set of isomorphism classes of pairs (E, P) where E is an elliptic curve and $P \in E$ a point of order N.

Let R be a ring and $b, c \in R$, then $E_{b,c}$ is the curve defined by

$$y^2 + (1-c)xy - by = x^3 - bx^2.$$

Define $R_{b,c} := \mathbb{Z}[b, c, \frac{1}{\Delta_{b,c}}]$ where $\Delta_{b,c}$ is the discriminant of the curve $E_{b,c}$ and define $Y := \operatorname{Spec} R_{b,c}$. The curve $E_{b,c}$ is an elliptic curve over Y and

$$P_0 := (0:0:1) \in E_{b,c}(Y).$$

Let $\Phi_N, \Psi_N, \Omega_N \in R_{b,c}$ be such that

$$(\Phi_N\Psi_N : \Omega_N : \Psi_N^3) = NP_0.$$

The equation $\Psi_N = 0$ is equivalent to P_0 having order dividing N. One can show that if $d \mid N$ then $\Psi_d \mid \Psi_N$. Define F_N by removing all factors coming form the Ψ_d with $d \mid N, d \neq N$ from Ψ_N, and

$$Y_1(N) := \operatorname{Spec} R_{b,c}[\tfrac{1}{N}]/F_N.$$

Let $\bar{b}, \bar{c} \in \operatorname{Spec} R_{b,c}[\frac{1}{N}]/F_N$ denote the equivalence classes of b, c and define $E_1(N) := E_{\bar{b},\bar{c}}$, it is an elliptic curve over $Y_1(N)$ and $P_1(N) := (0:0:1)$ is a point on it.

Let S be a scheme over $\mathbb{Z}[\frac{1}{N}]$ and $X \in Y_1(N)(S)$, then $E_1(N) \times_X S$ is an elliptic curve over R and the order of $P_1(N) \times_X S$ as well as that of all its geometric fibers is N. Conversely if E is an elliptic curve over S and $P \in E(R)[N]$ is such that the order of P is N in all geometric fibers, then there exist unique $b, c \in \mathcal{O}_R$ such that $F_N(b,c) = 0$ and $(E, P) \cong (E_{b,c}, P_0)$, furthermore this isomorphism is unique. So the pair b, c defines a point $X \in Y_1(N)(S)$ such that $(E, P) \cong (E_1(N)_S, P_1(N)_S)$.

We have seen that in the complex world the points of $\mathbf{Y}_1(N)$ correspond to pairs (E, P) where E is elliptic curves over $\mathbb{C}$ and P a point of order N. In the algebraic world we have seen that if R is a $\mathbb{Z}[\frac{1}{N}]$ algebra, then the points in $Y_1(N)(R)$ correspond to pairs (E, P) where E is an elliptic curve over R and $P \in E(R)[N]$

is such that the order of P is also N in all geometric fibers, in other words $Y_1(N)$ represents the functor which takes an R algebra to the set of isomorphism classes of pairs (E,P) of elliptic curve over R together with a point of order N. Taking $R=\mathbb{C}$ one obtains an isomorphism $\mathbf{Y}_1(N)\cong Y_1(N)(\mathbb{C})$ of Riemann surfaces. The curve $Y_1(N)$ is smooth over $\mathbb{Z}[\frac{1}{N}]$ and has geometrically irreducible fibers, see [Deligne and Rapoport(1975), Ch. IV].

2.1.1. *The universal elliptic curve with a point of order N.* In the above discussion we have seen that the pair $(E_1(N),P_1(N))$ is pair of an elliptic curve over $Y_1(N)$ together with a point $P_1(N)\in E_1(N)(Y_1(N))$ of order N all whose geometric fibers are also of order N. And we have even seen for R a $\mathbb{Z}[1/N]$-algebra that every pair (E,P) where E is an elliptic curve over R and $P\in E(R)$ a point of order N all whose geometric fibers also have order N can be obtained as the base change of $(E_1(N),P_1(N))$ along a unique morphism $X:\operatorname{Spec}R\to Y_1(N)$. The pair $(E_1(N),P_1(N))$ is called the universal elliptic curve with a point of order N. Now $(E_1(N)(\mathbb{C}),P_1(N)(\mathbb{C}))$ is a smooth family of elliptic curves with a smooth family of points of order N over $Y_1(N)(\mathbb{C})$ and this family can actually also be constructed directly in the complex world. Let $\mathbb{Z}^2$ act on $\mathbb{C}\times\mathbb{H}$ by $(m,n)(z,\tau)=(z+m\tau+n,\tau)$. Then the fiber of $(\mathbb{C}\times\mathbb{H})/\mathbb{Z}^2$ above $\tau\in\mathbb{H}$ is the elliptic curve E_τ, and the map $\mathbf{P}_1(N):\mathbb{H}\to(\mathbb{C}\times\mathbb{H})/\mathbb{Z}^2$ which sends τ to $(1/N \mod \mathbb{Z}\tau+\mathbb{Z},\tau)$ is a point of order N. If one lets $\mathrm{SL}_2(\mathbb{Z})$ act on $\mathbb{C}\times\mathbb{H}$ by

$$\begin{aligned}\mathrm{SL}_2(\mathbb{Z})\times(\mathbb{C}\times\mathbb{H})&\to\mathbb{C}\times\mathbb{H}\\ \left(\begin{bmatrix}a&b\\c&d\end{bmatrix},(z,\tau)\right)&\mapsto\left(\frac{z}{c\tau+d},\frac{a\tau+b}{c\tau+d}\right)\end{aligned}\tag{14}$$

Then one can make the semi-direct product $\mathbb{Z}^2\rtimes\mathrm{SL}_2(\mathbb{Z})$ act on $\mathbb{C}\times\mathbb{H}$ by

$$((m,n),\gamma)(z,\tau)=(m,n)(\gamma(z,\tau)).$$

Now define $\mathbf{E}_1(N):=(\mathbb{Z}^2\rtimes\Gamma_1(N))\backslash(\mathbb{C}\times\mathbb{H})$. The map $\mathbf{E}_1(N)\to\mathbf{Y}_1(N)$ which sends $(\mathbb{Z}^2\rtimes\Gamma_1(N))(z,\tau)$ to $\Gamma_1(N)\tau$ makes $\mathbf{E}_1(N)$ into a family of curves over $\mathbf{Y}_1(N)$. Using $N>4$ one sees that the stabilizer of τ in $\Gamma_1(N)$ is trivial for all $\tau\in\mathbb{H}$. This triviality of the stabilizers implies that the fiber of $\mathbf{E}_1(N)\to\mathbf{Y}_1(N)$ above $\Gamma_1(N)\tau$ is isomorphic to E_τ for all $\tau\in\mathbb{H}$. One checks that the map $\mathbf{P}_1(N):\mathbb{H}\to(\mathbb{C}\times\mathbb{H})/\mathbb{Z}^2$ induces a map $\mathbf{P}_1(N):\mathbf{Y}_1(N)\to\mathbf{E}_1(N)$ by taking the quotient by $\Gamma_1(N)$ on both sides. The pair $(\mathbf{E}_1(N),\mathbf{P}_1(N))$ is the universal elliptic curve with a point of order N in the complex setting. And it is isomorphic to $(E_1(N)(\mathbb{C}),P_1(N)(\mathbb{C}))$.

2.2. **The modular curve $X_1(N)$.** The curve $\mathbf{Y}_1(N)$ of the previous section is not compact and the curve $Y_1(N)$ is not proper over $\mathbb{Z}[\frac{1}{N}]$. But compactness and properness are properties that are useful for studying curves (and higher dimensional varieties/schemes). The modular curves $\mathbf{X}_1(N)$, respectively $X_1(N)$ that will be defined in this section are compact, respectively proper over $\mathbb{Z}[\frac{1}{N}]$. The curves $\mathbf{Y}_1(N)$ respectively $Y_1(N)$ will be open and dense parts of them.

The j-invariant induces a holomorphic map $j : \mathbf{Y}_1(N) \to \mathbb{C}$ by sending (E, P) to $j(E)$. This turns $\mathbf{Y}_1(N)$ into a finite ramified cover of $\mathbb{C}$. See $\mathbb{C}$ as an open in $\mathbb{P}^1(\mathbb{C})$ whose complement is the point ∞. Take $D \subseteq \mathbb{P}^1(\mathbb{C})$ a punctured disc centered at ∞. By choosing D small enough one can assure that $j^{-1}(D)$ is a disjoint union of punctured discs. The Riemann surface $\mathbf{X}_1(N)$ is the Riemann surface obtained from the Riemann surface $\mathbf{Y}_1(N)$ by filling the holes in these punctured discs. The map j turns $\mathbf{X}_1(N)$ into a finite ramified cover of $\mathbb{P}^1(\mathbb{C})$. One has $j^{-1}(\infty) = \mathbf{X}_1(N) \setminus \mathbf{Y}_1(N)$. The set $j^{-1}(\infty)$ is a finite set and its elements are called the cusps.

The j-invariant induces a morphism of $\mathbb{Z}[\frac{1}{N}]$-schemes $j : Y_1(N) \to \mathbb{A}^1_{\mathbb{Z}[1/N]}$ which sends $(E, P) \in Y_1(N)(T)$ to $j(E) \in \mathcal{O}_T$ for all $\mathbb{Z}[\frac{1}{N}]$-schemes T. See $\mathbb{A}^1_{\mathbb{Z}[1/N]}$ as an open subscheme of $\mathbb{P}^1_{\mathbb{Z}[1/N]}$ whose complement is the closed subscheme ∞. The generic point of $\mathbb{P}^1_{\mathbb{Z}[1/N]}$ is $\operatorname{Spec}\mathbb{Q}(j)$ and by viewing j as element of $\mathbb{Q}(Y_1(N))$ we see that $\mathbb{Q}(j) \subseteq \mathbb{Q}(Y_1(N))$ is a finite extension of fields. The curve $X_1(N)$ is defined as the normalization of $\mathbb{P}^1_{\mathbb{Z}[1/N]}$ in $\mathbb{Q}(Y_1(N))$. The map j turns $X_1(N)$ into a finite ramified cover of $\mathbb{P}^1_{\mathbb{Z}[1/N]}$. One has $j^{-1}(\infty) = X_1(N) \setminus Y_1(N)$. The scheme $j^{-1}(\infty)_{\mathbb{Z}[1/N,\zeta_N]\cap\mathbb{R}}$ is a disjoint union of copies of $\operatorname{Spec}\mathbb{Z}[\frac{1}{N}, \zeta_N] \cap \mathbb{R}$.

In the complex world there is also a second way to construct the underlying topological space of the Riemann surface $\mathbf{X}_1(N)$. For this one first defines $\mathbb{H}^* := \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ and one extends the action of $\mathrm{SL}_2(\mathbb{Z})$ to $\mathbb{H}^*$ still using the formula $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \tau = \frac{a\tau+b}{c\tau+d}$, where one defines $\frac{a\infty+b}{c\infty+d} = \frac{a}{c}$ and $\frac{az+b}{cz+d} = \infty$ if $cz + d = 0$. One can show that $\mathrm{SL}_2\,\mathbb{Z}$ acts transitively on $\mathbb{P}^1(\mathbb{Q})$. One topologizes $\mathbb{H}^*$ by saying that $\mathbb{H}_{\mathrm{im}\,Z>x} \cup \{\infty\}$ with $x \in \mathbb{R}_{>0}$ forms a basis of open neighbourhoods of ∞ and requiring that the topology is invariant under the action of $\mathrm{SL}_2(\mathbb{Z})$. One can show that there is a unique isomorphism of topological spaces between $\mathbf{X}_1(N)$ and $\Gamma_1(N)\backslash\mathbb{H}^*$ that is the identity on $Y_1(N) := \Gamma_1(N)\backslash\mathbb{H}$. This allows one to identify the cusps of $\mathbf{X}_1(N)$ with $\Gamma_1(N)\backslash\mathbb{P}^1(\mathbb{Q})$.

2.2.1. *Moduli interpretation of the cusps.* In Section 2.1 we saw that there exists an elliptic curve $E_1(N)$ over $Y_1(N)$ which has a point of order N that also has order N in all geometric fibers, and that for every $\mathbb{Z}[\frac{1}{N}]$ algebra R every elliptic curve over R together with a point of order N that also has order N in all geometric fibers is the base change of $E_1(N)$ to R for a unique morphism $X : \operatorname{Spec} R \to Y_1(N)$. Using the notion of generalized elliptic curve this story extends to $X_1(N)$. There is a unique extension $(E_1'(N), P_1'(N))$ of the pair $(E_1(N), P_1(N))$ over $Y_1(N)$ to $X_1(N)$ such that the geometric fibers of $E_1'(N)$ over $X_1(N)$ are generalized elliptic curves, the point $P_1'(N)$ lies in the smooth locus of $E_1'(N)$, the geometric fibers of $P_1'(N)$ are all points of order N and for each geometric fiber $P_1'(N)$ is a generator of the component group.

Theorem 2.1. *[Deligne and Rapoport(1975), Ch. IV] This pair $(E_1'(N), P_1'(N))$ mentioned above is universal, meaning that if S is a scheme over $\mathbb{Z}[\frac{1}{N}]$ and (E, P) is a pair where E is an elliptic curve over S and $P \in E(S)$ a point of order N such* *that all the geometric fibers of P are also of order N and generate the component group of their fiber, then there exists a unique $X : S \to X_1(N)$ such that (E, P) is isomorphic to the base change of $(E_1'(N), P_1'(N))$.*

In particular, if K is an algebraically closed field and $s \in (X_1(N) \setminus Y_1(N))(K)$, then $E_1'(N)_s$ is a Néron d-gon for some integer d and $P_1'(N)_s$ is a point of order N that generates the component group of the Néron d-gon. Since the component group of a Néron d-gon is $\mathbb{Z}/d\mathbb{Z}$ this means that $d \mid N$.

The Tate curve given by Eq. (13) gives a way to study the curve $E_1'(N)$ over $X_1(N)$ in the neighbourhood of the cusps. Let $d \mid N$ be an integer and denote by $E_{q,d}$ the base change of E_q to $\mathbb{Z}[[q^{1/d}]]$. The scheme $E_{q,d}$ is not smooth over $\mathbb{Z}[[q^{1/d}]]$, but if $d = 1$ then it is at least still a regular scheme. If $d > 1$ then the singularities of $E_{q,d}$ can be resolved by blowing up the point $(q, x', y') = (0, 0, 0)$ exactly $\lfloor \frac{d}{2} \rfloor$ times, let $\tilde{E}_{q,d}$ denote the resulting scheme, its fiber over $q^{1/d} = 0$ is the Néron d-gon over $\mathbb{Z}$, and for every field K one has that $\tilde{E}_{q,d,K[[q^{1/d}]]}$ is the minimal regular model of $E_{q,d,K[[q^{1/d}]]}$. Consider x' and y' of Eq. (13) as elements of $\mathbb{Z}((u))[[q]]$ and let i, j be two integers. Evaluating x' and y' at $u = q^i \zeta_N^j$ gives a $\mathbb{Z}[\frac{1}{N}, \zeta_N][[q^{1/d}]]$ point of $\tilde{E}_{q,d}$, which we will denote by $P_{d,i,j}$. This point lies in the smooth locus and its order is a divisor of N. Actually the map

$$\alpha : \mathbb{Z}/d\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z} \to \tilde{E}_{q,d}^{sm}(\mathbb{Z}[\tfrac{1}{N}, \zeta_N][[q^{\frac{1}{d}}]]) \tag{15}$$

$$i, j \mapsto P_{d,i,j} \tag{16}$$

is a well defined injective group homomorphism. The point $\alpha(1, 0)$ is a generator of the component group at $q^{1/d} = 0$ and $\alpha(0, 1)$ lies in the identity component. Define $A_d \subset \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/d\mathbb{Z}$ to be the set of elements of order N whose first coordinate generates $\mathbb{Z}/d\mathbb{Z}$. The set A_d is exactly the set of (i, j) such that the pair $(\tilde{E}_{q,d}, \alpha(i, j))$ gives a point $s_{d,i,j} \in X_1(N)(\mathbb{Z}[\frac{1}{N}, \zeta_N][[q^{1/d}]])$. Let $s'_{d,i,j} \in X_1(N)(\mathbb{Z}[\zeta_N])$ be the point obtained by setting $q^{1/d} = 0$, then $s'_{d,i,j}$ is a cusp, and the map $s_{d,i,j} : \operatorname{Spec} \mathbb{Z}[\zeta_N][[q^{1/d}]] \to X_1(N)$ induces an isomorphism between $\mathbb{Z}[\zeta_N][[q^{1/d}]]$ and the completion of $X_1(N)_{\mathbb{Z}[\frac{1}{N}, \zeta_N]}$ along $s'_{d,i,j}$. Every Néron d-gon together with a point of order N that generates the component group is obtained from some $s'_{d,i,j}$ with $(i, j) \in A_d$, showing that $\left\{s'_{d,i,j} | d \mid N, (i, j) \in A_d\right\}$ is exactly the set of cusps of $X_1(N)_{\mathbb{Z}[\frac{1}{N}\zeta_N]}$, however two different elements of A_d might give the same cusp, indeed one can make $\mu_d(\mathbb{Z}[\frac{1}{N}, \zeta_d]) \times \{\pm 1\}$ act on A_d by $\zeta_d(i, j) = (i, j + iN/d)$ and $-(i, j) = (-i, -j)$. This action is compatible with the action of $\mu_d \times \{\pm 1\}$ on the set of points of order N of the Néron d-gon, showing that $s'_{d_1,i_1,j_1} = s'_{d_2,i_2,j_2}$ if and only if $d_1 = d_2$ and (i_1, j_1) and (i_2, j_2) are in the same orbit under this action.

3. Modular forms

Let $k > 0$ be an integer, $f : \mathbb{H} \to \mathbb{C}$ be a holomorphic function and $\gamma = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \in \mathrm{SL}_2(\mathbb{Z})$, define $f[\gamma]_k : \mathbb{H} \to \mathbb{C}$ to be the function given by $f[\gamma]_k(\tau) := (c\tau + d)^{-k} f(\gamma\tau)$. The map $f \to f[\gamma]_k$ defines a right action of $\mathrm{SL}_2(\mathbb{Z})$ on the set of all holomorphic functions $\mathbb{H} \to \mathbb{C}$ called the weight k action.

Definition 3.1. Let $k > 0$ be an integer and $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ be a finite index subgroup, then a modular form of weight k for Γ is a continuous function $f : \mathbb{H}^* \to \mathbb{C}$ such that:

(1) f is invariant under the weight k action of Γ, i.e. $f = f[\gamma]_k$ for all $\gamma \in \Gamma$.
(2) f is holomorphic when restricted to $\mathbb{H}$.

The function f is called a cusp form if $f(x) = 0$ for all $x \in \mathbb{P}^1(\mathbb{Q})$.

Where one should note that in this definition f is required to be continuous on all of $\mathbb{H}^*$. If one instead just requires f to be continuous on $\mathbb{H}$ one needs to add an extra condition that is called being "holomorphic at the cusps". This complex analytic definition of modular forms does not carry over to the algebraic world, however it can be reinterpreted in a way that does make sense in the algebraic world. Namely one can define $\omega_{\Gamma,k}$ to be the sheaf on $X(\Gamma) := \Gamma\backslash\mathbb{H}^*$ whose functions on $\Gamma\backslash U$ are the continuous functions $f : U \to \mathbb{C}$ invariant under the weight k action of Γ that are holomorphic when restricted to $\mathbb{H} \cap U$ for all open $U \subset \mathbb{H}^*$ that are invariant under Γ. If either k is even or Γ acts freely on $\mathbb{H}$ then the sheaf $\omega_{\Gamma,k}$ is a line bundle on $X(\Gamma)$, i.e. it is a sheaf of $\mathcal{O}_{X(\Gamma)}$ modules that is locally free of rank 1. The global sections of $\omega_{\Gamma,k}$ are exactly the modular forms of weight k. This line bundle $\omega_{\Gamma,k}$ is the object that does generalize to the algebraic world, at least if one requires that Γ is a congruence subgroup:

Definition 3.2. Let N be an integer and define

$$\Gamma(N) := \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \mod N \right\}.$$

A *congruence subgroup* is a subgroup $\Gamma \subseteq \mathrm{SL}_2(\mathbb{Z})$ such that there exists an integer N for which $\Gamma(N) \subseteq \Gamma$.

For simplicity we will restrict ourselves to congruence subgroups Γ that contain $\Gamma_1(N)$ as a normal subgroup for some N in the discussion below. First we will discuss modular forms of weight k for $\Gamma_1(N)$ with $N > 4$ and only later will we discuss it for its groups that contain $\Gamma_1(N)$.

3.1. Modular forms for $\Gamma_1(N)$.

Let $N > 4$ be an integer. Over the curve $\mathbf{X}_1(N)$ we have the universal curve $\mathbf{E}_1(N)$, and we have the zero section $0 : \mathbf{X}_1(N) \to \mathbf{E}_1(N)$. This means we can look at the sheaf $\Omega^1_{\mathbf{E}_1(N)/\mathbf{X}_1(N)}$ of relative differential forms on $\mathbf{E}_1(N)$, this sheaf is locally free of rank 1 when restricted to the locus of

$\mathbf{E}_1(N)$ where it is smooth over $\mathbf{X}_1(N)$. Define

$$\omega_{\mathbf{E}_1(N)/\mathbf{X}_1(N)} := 0^*\Omega^1_{\mathbf{E}_1(N)/\mathbf{X}_1(N)}.$$

One has that $\omega_{\mathbf{E}_1(N)/\mathbf{X}_1(N)} \cong \omega_{\Gamma_1(N),1}$ and more generally $\omega^{\otimes k}_{\mathbf{E}_1(N)/\mathbf{X}_1(N)} \cong \omega_{\Gamma_1(N),k}$. Indeed, let $\pi : \mathbb{H}^* \to X_1(N)$ be the quotient map, then $\pi^*\omega_{\omega_{\mathbf{E}_1(N)/\mathbf{X}_1(N)}}$ is a free sheaf of rank 1 when restricted to $\mathbb{H}$. This is because one has $\pi^*\omega_{\Gamma_1(N)} \cong 0^*\Omega^1_{((\mathbb{C}\times\mathbb{H})/\mathbb{Z}^2))/\mathbb{H})}$ and the latter is generated by dz where z is the coordinate on $\mathbb{C}$. Since

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mathrm{d}z = \mathrm{d}\frac{z}{c\tau + d} = \frac{1}{c\tau + d}\mathrm{d}z$$

it follows that $f \mapsto 2\pi i f \mathrm{d}z = f\frac{\mathrm{d}u}{u}$ gives an isomorphism between $\omega^{\otimes k}_{\mathbf{E}_1(N)/\mathbf{X}_1(N)}$ and $\omega_{\Gamma_1(N),k}$ on $\mathbf{Y_1(N)}$. Using the Tate curve over $\mathbb{C}$ one can show that $2\pi i \mathrm{d}z = \frac{\mathrm{d}u}{u}$ is also a generator of $\Omega^1_{\mathbf{E}_1(N)/\mathbf{X}_1(N)}$ in a neighbourhood of the 0 section at the cusps ($u = 1$ at the zero section), hence the isomorphism over $\mathbf{Y}_1(N)$ extends to one over $\mathbf{X}_1(N)$.

In the previous section it was already shown that modular forms of weight k for $\Gamma_1(N)$ can be seen as sections of $\omega_{\Gamma_1(N),k}$ and using the isomorphism $\omega^{\otimes k}_{\mathbf{E}_1(N)/\mathbf{X}_1(N)} \cong \omega_{\Gamma_1(N),k}$ one can even see them as sections of $\omega^{\otimes k}_{\mathbf{E}_1(N)/\mathbf{X}_1(N)}$. This last definition is the definition that carries over to the algebraic world.

Definition 3.3. Let $N > 4$ and k be integers and R a $\mathbb{Z}[\frac{1}{N}]$ algebra. Define

$$\omega_{X_1(N),R,k} := \left(0^*\Omega^1_{E_1(N)_R/X_1(N)_R}\right)^{\otimes k}.$$

An *R valued modular form of weight k for $X_1(N)$* is a global section f of $\omega_{X_1(N),R,k}$. A modular form f is called a cusp form if it has zeros at all cusps, i.e. it is zero on $X_1(N)_R \setminus Y_1(N)_R$.

The above discussion shows that if one takes $R = \mathbb{C}$ then this definition agrees with the complex analytic definition.

3.2. **Modular forms in weight 2.** In weight 2 there is even a different interpretation of modular forms. The reason for this is that

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \mathrm{d}\tau = \mathrm{d}\frac{a\tau + b}{c\tau + d} = \frac{a(c\tau + d) - c(a\tau + b)}{(c\tau + d)^2}\mathrm{d}\tau = \frac{1}{(c\tau + d)^2}\mathrm{d}\tau,$$

showing that if f is a complex analytic modular form of weight 2 for some congruence subgroup Γ, then $2\pi i f \mathrm{d}\tau = f\frac{\mathrm{d}q}{q}$ is a differential on $\mathbb{H}$ that is invariant under the action of Γ. In particular, $f\frac{\mathrm{d}q}{q}$ descends to a differential on $Y_1(N)$. Using the description of the formal neighbourhoods of the cusps one can show that this differential has no poles at the cusps if and only if f is a cusp form, so that

$f \mapsto f\frac{\mathrm{d}q}{q}$ gives an isomorphism $\omega_{\Gamma_1(N),2} \cong \Omega^1_{\mathbf{X}_1(N)/\mathbb{C}}(cusps)$ called the Kodaira-Spencer isomorphism. This isomorphism extends to the algebraic world as an isomorphism $\omega_{X_1(N),\mathbb{Z}[1/N],2} \cong \Omega^1_{X_1(N)/\mathbb{Z}[1/N]}(cusps)$, where the isomorphism is given by $(\frac{\mathrm{d}u}{u})^{\otimes 2} \mapsto \frac{\mathrm{d}q}{q}$ at the Tate curve.

This discussion shows cusp forms of weight two for $X_1(N)$ over a ring R can be interpreted as global sections of $\Omega^1_{\mathbf{X}_1(N)/R}$.

4. The modular curves $X_0(N)$ and $X_\mu(N)$.

In the sections on $Y_1(N)$ and $X_1(N)$ we saw that these curves parametrize elliptic curves together with a point of order N. The curves $Y_0(N)$ and $X_0(N)$ are the curves that parametrize elliptic curves together with a cyclic subgroup of order N.

The complex setting will be described first. Define

$$\Gamma_0(N) := \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \mid c \equiv 0 \mod N \right\},$$

and recall that if $\tau \in \mathbb{H}$, then E_τ denotes the curve $\mathbb{C}/(\tau\mathbb{Z} + \mathbb{Z})$. In the discussion on $\mathbf{Y}_1(N)$ it was shown that if $\tau_1, \tau_2 \in \mathbb{H}$ then pairs $(E_{\tau_1}, 1/N)$ and $(E_{\tau_2}, 1/N)$ are isomorphic if and only if there exists a γ in $\Gamma_1(N)$ such that $\tau_2 = \gamma\tau_1$. Similarly one can show, replacing the point $1/N \in E_\tau$ by the subgroup generated by $1/N$, that $(E_{\tau_1}, \langle 1/N \rangle)$ and $(E_{\tau_2}, \langle 1/N \rangle)$ are isomorphic if and only if there exists a $\gamma \in \Gamma_0(N)$ such that $\tau_2 = \gamma\tau_1$. This shows that over $\mathbb{C}$ the isomorphism classes of pairs (E, G) of elliptic curve together with a cyclic subgroup of order N are in one to one correspondence with $\Gamma_0\backslash\mathbb{H}$. So the modular curve $\mathbf{Y}_0(N)$ is defined to be $\Gamma_0\backslash\mathbb{H}$. One can compactify $\mathbf{Y}_0(N)$ in a similar way to $\mathbf{Y}_1(N)$ and the resulting compactification will be denoted by $\mathbf{X}_0(N) = \Gamma_0\backslash\mathbb{H}^*$.

Note that $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ so that we could also have constructed $\mathbf{Y}_0(N)$ and $\mathbf{X}_0(N)$ as quotients of $\mathbf{Y}_1(N)$ and $\mathbf{X}_1(N)$ by $\Gamma_0(N)/\Gamma_1(N)$. The map $\Gamma_0(N) \to (\mathbb{Z}/N\mathbb{Z})^*$ given by $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right] \mapsto d$ is a surjective group homomorphism whose kernel is $\Gamma_1(N)$ showing that $\Gamma_0(N)/\Gamma_1(N) \cong (\mathbb{Z}/N\mathbb{Z})^*$. One can even interpret the action of $(\mathbb{Z}/N\mathbb{Z})^*$ on $\mathbf{X}_1(N)$ directly, since $d \in (\mathbb{Z}/N\mathbb{Z})^*$ corresponds to sending the pair (E, P) of elliptic curve with point of order N to (E, dP). The automorphism of $X_1(N)$ corresponding to $d \in (\mathbb{Z}/N\mathbb{Z})^*$ is denoted by $\langle d \rangle$ and is called a *diamond operator*.

The action of $d \in (\mathbb{Z}/N\mathbb{Z})^*$ given by by $(E, P) \mapsto (E, dP)$ also makes sense in the algebraic world and gives an action on the $\mathbb{Z}[\frac{1}{N}]$-schemes $Y_1(N)$ and $X_1(N)$. One uses this action to define the modular curves $Y_0(N)$ resp. $X_0(N)$ to be $Y_1(N)/(\mathbb{Z}/N\mathbb{Z})^*$ resp. $X_1(N)/(\mathbb{Z}/N\mathbb{Z})^*$. We saw that $Y_1(N)(R)$ can be identified with the set of isomorphism classes of pairs (E, P) of elliptic curve together over R with a point of order N for all $\mathbb{Z}[\frac{1}{N}]$-algebras R. However for $Y_0(N)$ this property fails. A pair (E, G) of elliptic curve over R together with a cyclic subgroup of order N still gives rise to an R valued point on $Y_0(N)$, but non isomorphic pairs (E_1, G_1) and (E_2, G_2) might give the same R point of $Y_0(N)$. Although in the case $R = \overline{K}$ is an algebraically closed field then $Y_0(N)(R)$ can still be identified with the set of isomorphism classes of elliptic curves with a point of order N, as we already saw over $\mathbb{C}$. An additional problem with $\mathbf{Y}_0(N)$ and $Y_0(N)$ is that there is no universal elliptic curve over them. If one tries to define the universal elliptic curve $\mathbf{E}_0(N) := (\mathbb{Z}^2 \rtimes \Gamma_0(N))\backslash(\mathbb{C} \times \mathbb{H})$ over $\mathbf{Y}_0(N)$, similar to what was done for $\mathbf{E}_1(N)$, then one runs into problems. This definition would still give a curve over $\mathbf{Y}_0(N) := \Gamma_0(N)\backslash\mathbb{H}$, however the proof that the fiber of $\mathbf{E}_1(N)$ over $\Gamma_1(N)\tau \in \mathbf{X}_1(N)$ is isomorphic to E_τ uses that $\Gamma_1(N)$ acts freely on $\mathbb{H}$ under the assumption $N > 4$. This is no longer true for $\Gamma_0(N)$, in fact since $-\operatorname{Id} := \left[\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right] \in \Gamma_0(N)$ and $-\operatorname{Id}$ acts trivially on $\mathbb{H}$ we see that the fiber of $\mathbf{E}_0(N)$ above $\Gamma_0(N)\tau$ is a quotient of $E_\tau/\pm 1$ which is not an elliptic curve, but something isomorphic to $\mathbb{P}^1(\mathbb{C})$. One has similar problems with trying to construct the universal elliptic curve over $X_0(N)$,

4.1. **Modular forms on $X_0(N)$.** The complex analytic definition of a modular form in Definition 3.1 is general enough to also work if one takes $\Gamma = \Gamma_0(N)$. However the algebraic definition 3.3 for modular forms on $X_1(N)$ uses the existence of the universal elliptic curve $E_1(N)$ over $X_1(N)$. This leads to problems when trying to define modular forms on $X_0(N)$, since we saw previously that we have no universal elliptic curve in this case. However these problems can by solved. Namely let R be a $\mathbb{Z}[\frac{1}{N}]$-algebra and let $\pi : X_1(N) \to X_0(N)$ denote the quotient map, then $\pi_*\omega_{X_1(N),R,k}$ is a sheaf on $X_0(N)$ with an action of $(\mathbb{Z}/N\mathbb{Z})^*$, taking $(\mathbb{Z}/N\mathbb{Z})^*$ invariants gives the desired sheaf of $X_0(N)$.

Definition 4.1. Let $N > 4$ and k be integers and R a $\mathbb{Z}[\frac{1}{N}]$ algebra. Define

$$\omega_{X_0(N),R,k} := (\pi_*\omega_{X_1(N),R,k})^{(\mathbb{Z}/N\mathbb{Z})^*}.$$

An *R valued modular form of weight k for $X_0(N)$* is a global section f of $\omega_{X_1(N),R,k}$. A modular form f is called a cusp form if it has zeros at all cusps, i.e. it is zero on $X_0(N)_R \setminus Y_0(N)_R$.

Let R be a flat $\mathbb{Z}[\frac{1}{N}]$-algebra, then $\Omega^1_{X_0(N)/R} \cong (\pi_*\Omega^1_{X_1(N)/R})^{(\mathbb{Z}/N\mathbb{Z})^*}$. This means that the Kodaira-Spencer isomorphism $\omega_{X_1(N),\mathbb{Z}[1/N],2} \cong \Omega^1_{X_1(N)/\mathbb{Z}[1/N]}(cusps)$ descends to an isomorphism $\omega_{X_0(N),\mathbb{Z}[1/N],2} \cong \Omega^1_{X_0(N)/\mathbb{Z}[1/N]}(cusps)$, showing that one can still see cusp forms over R as one forms. However if R is a ring that is not flat over $\mathbb{Z}[\frac{1}{N}]$ there are some troubles that can arise, especially rings of characteristic 2 and 3 pose problems. More details on different ways of viewing cusp forms as differential forms and the difficulties that arise in characteristics 2 and 3 can be found in [Mazur(1977), §II.4].

4.2. **The modular curve $X_\mu(N)$..** The modular curve $X_\mu(N)$ is just a slight variation on the modular curve $X_1(N)$. The curve $X_1(N)$ parametrizes pairs (E, P) of an elliptic curve together with a point of order N, or equivalently pairs (E, α) where $\alpha : \mathbb{Z}/N\mathbb{Z} \to E$ is a closed immersion of the constant group scheme into

E. The modular curve $X_\mu(N)$ parametrizes pairs (E,β) where $\beta : \mu_N \to E$ is a closed immersion of the group of N-th roots of unity into E. Since over $\mathbb{Z}[\frac{1}{N},\zeta_N]$ one has $\mu_N \cong \mathbb{Z}/N\mathbb{Z}$, one sees that $X_1(N)_{\mathbb{Z}[1/N,\zeta_N]} \cong X_\mu(N)_{\mathbb{Z}[1/N,\zeta_N]}$. This isomorphism shows in particular that $X_1(N)$ and $X_\mu(N)$ are isomorphic over all algebraically closed fields and that $X_\mu(N)$ and $X_1(N)$ are twists of each other over $\mathbb{Z}[\frac{1}{N},\zeta_N]$. Since $\zeta_N \in \mathbb{C}$ there is nothing that really changes in the complex world so that we still can see $X_\mu(N)(\mathbb{C})$ as $\mathbf{X}_1(N) = X_1(N)(\mathbb{C})$. However over rings not containing ζ_N there is a difference. The twisting of $X_1(N)$ that gives $X_\mu(N)$ can even be made explicit by the isomorphism

$$X_\mu(N) \cong \left(X_1(N) \times_{\mathbb{Z}[\frac{1}{N}]} \mathbb{Z}\left[\frac{1}{N},\zeta_N\right]\right)/(\mathbb{Z}/N\mathbb{Z})^*,$$

where $d \in (\mathbb{Z}/N\mathbb{Z})^*$ acts on $X_1(N)$ via the diamond operator $\langle d\rangle$ and on $\mathbb{Z}[\frac{1}{N},\zeta_N]$ via $\zeta_N \mapsto \zeta_N^d$. In contrast to $X_0(N)$, the modular curve $X_\mu(N)$ does have a universal elliptic curve over it. This universal elliptic curve is denoted by $E_\mu(N)$ and the entire story about modular forms on $X_1(N)$ translates directly to a description of the modular forms on $X_\mu(N)$. For a $\mathbb{Z}[\frac{1}{N}]$ algebra R one can define

$$\omega_{X_\mu(N),R,k} := \left(0^*\Omega^1_{E_\mu(N)_R/X_\mu(N)_R}\right)^{\otimes k}.$$

similar to Definition 3.3, and say that an R-valued modular form of weight k on $X_\mu(N)$ is a global section of $\omega_{X_\mu(N),R,k}$. Also the Kodaira-Spencer isomorphism $\omega_{X_\mu(N),\mathbb{Z}[1/N],2} \cong \Omega^1_{X_\mu(N)/\mathbb{Z}[1/N]}(cusps)$ continues to exist.

References

[Deligne and Rapoport(1975)] P. Deligne and M. Rapoport. Correction to: "Les schémas de modules de courbes elliptiques" (*modular functions of one variable, ii* (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972), pp. 143–316, Lecture Notes in Math., Vol. 349, Springer, Berlin, 1973). In *Modular functions of one variable, IV (Proc. Internat. Summer School, Univ. Antwerp, Antwerp, 1972)*, pages p. 149. Lecture Notes in Math., Vol. 476. Springer, Berlin, 1975.

[Mazur(1977)] B. Mazur. Modular curves and the Eisenstein ideal. *Inst. Hautes Études Sci. Publ. Math.*, (47):33–186 (1978), 1977. ISSN 0073-8301. URL http://www.numdam.org/item?id=PMIHES_1977__47__33_0.

[Silverman(1994)] Joseph H. Silverman. *Advanced topics in the arithmetic of elliptic curves*, volume 151 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1994. ISBN 0-387-94328-5. doi: 10.1007/978-1-4612-0851-8. URL http://dx.doi.org/10.1007/978-1-4612-0851-8.