MODULAR CURVES AND HEEGNER POINTS

ALEXANDRU GHITZA

Abstract. The ultimate goal of these lectures is to define Heegner points, which are the building blocks for Kolyvagin's Euler system. We start by defining the modular curves $X_0(N)$ and proving that j and j_N generate their field of rational functions. We use the modular equation to give a model of $X_0(N)$ over $\mathbb{Q}$, then discuss parametrizations of elliptic curves by modular curves. We give a moduli space interpretation for $X_0(N)$, define Heegner points and discuss their basic properties.

From the paper ([Gro91]), the construction of y_n (p. 238, from beginning of section 3).

References: Cox ([Cox89]), Milne ([Mil]), Rohrlich ([Roh97]).

1. Modular curves

Let $\mathbb{H}$ denote the upper half plane $\{z \in \mathbb{C} : \mathrm{Im}(z) > 0\}$ and let N be a positive integer. Consider the following subgroup of $SL_2(\mathbb{Z})$:

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) : c \equiv 0 \pmod N \right\}.$$

$\Gamma_0(N)$ acts on $\mathbb{H}$ via fractional linear transformations:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d},$$

and we consider the quotient space $Y_0(N) = \mathbb{H}/\Gamma_0(N)$. This is a non-compact Riemann surface. Similarly, let $\mathbb{H}^* = \mathbb{H} \cup \mathbb{P}^1(\mathbb{Q})$ denote the extended upper half plane. $\Gamma_0(N)$ acts on $\mathbb{H}^*$ in the same way, and the quotient is a compact Riemann surface, the modular curve $X_0(N)$. The finitely many elements of $X_0(N) \setminus Y_0(N)$ are called cusps.

A modular function for $\Gamma_0(N)$ is a meromorphic function on $X_0(N)$. Equivalently, it is a meromorphic function on $\mathbb{H}$ satisfying

1. $f(\gamma z) = f(z)$ for all $\gamma \in \Gamma_0(N), z \in \mathbb{H}$;
2. $f(z)$ meromorphic at the cusps.

The second condition needs some explanation. We start by considering the cusp ∞. $f(z)$ is invariant under $T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, so $f(z+1) = f(z)$. Therefore $f(z)$ can be expressed as a function $f^*(q)$ of the variable $q = e^{2\pi i z}$. As z ranges over $\mathbb{H}$, $q(z)$ ranges over a punctured disk $0 < |q| < \rho$. We say that $f(z)$ is meromorphic at ∞ if $f^*(q)$ is meromorphic at $q = 0$, that is if f has an expansion

$$f(z) = \sum_{n=n_0}^{\infty} a_n q^n.$$

Now if $\tau \neq \infty$ is a cusp, there exists $\gamma \in SL_2(\mathbb{Z})$ such that $\tau = \gamma(\infty)$. The function $z \mapsto f(\gamma z)$ is invariant under $\gamma\Gamma_0(N)\gamma^{-1}$, and $f(\gamma z)$ is required to be meromorphic at ∞.

A modular form for $\Gamma_0(N)$ of weight $2k$ is a holomorphic function on $\mathbb{H}$ such that

1. $f(\gamma z) = (cz+d)^{2k} f(z)$ for all $\gamma \in \Gamma_0(N), z \in \mathbb{H}$;
2. f is holomorphic at the cusps.

A modular form is called a cusp form if it is zero at the cusps.

We define

$$g_2(z) = 60 \sum_{(m,n)\neq(0,0)} \frac{1}{(m+nz)^4}$$

Kolyvagin seminar, February 15, 2000.

$$g_3(z) = 140 \sum_{(m,n)\neq(0,0)} \frac{1}{(m+nz)^6}$$

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

$$j(z) = 1728\frac{g_2(z)^3}{\Delta(z)}.$$

g_2, resp. g_3 are modular forms of weights 4, resp. 6. Δ is a cusp form of weight 12. j is a modular function. The q-expansion of $j(z)$ is

$$j(z) = \frac{1}{q} + \sum_{n=0}^{\infty} c_n q^n,$$

where $c_n \in \mathbb{Z}$ for all n.

2. The modular equation

Theorem 1. j is holomorphic on $\mathbb{H}$. $j'(z) \neq 0$ for all $z \in \mathbb{H} \setminus \{\gamma i, \gamma e^{2\pi i/3} : \gamma \in SL_2(\mathbb{Z})\}$. If $z = \gamma i$, then $j(z) = 1728$, $j'(z) = 0$ and $j''(z) \neq 0$. If $z = \gamma e^{2\pi i/3}$, then $j(z) = j'(z) = j''(z) = 0$ and $j'''(z) \neq 0$.

Proof. See [Cox89], pp. 221–222. □

Let $j_N(z) = j(Nz)$ for all $z \in \mathbb{H}^*$.

Theorem 2. $j_N(z)$ is a modular function for $\Gamma_0(N)$.

Proof. See [Cox89], pp. 226–229. □

Lemma. For all $z \in \mathbb{H}$ there exists $\gamma \in SL_2(\mathbb{Z})$ such that $|\operatorname{Re}(\gamma z)| \leq 1/2$ and $|\operatorname{Im}(\gamma z)| \geq 1/2$.

Proof. See [Cox89], p. 222. □

Note that a holomorphic modular function f for $SL_2(\mathbb{Z})$ is a polynomial in $j(z)$. f is meromorphic at ∞, so its q-expansion has finitely many terms in q^{-1}. Since the q-expansion for j starts with a q^{-1}, there exists a polynomial $P(X) \in \mathbb{C}[X]$ such that $g(z) = f(z) - P(j(z))$ is holomorphic at ∞. But then $g(z)$ is holomorphic on $X_0(N)$, which is a compact Riemann surface. Therefore g is a constant and $f(z)$ is a polynomial in $j(z)$.

Theorem 3. Every modular function for $SL_2(\mathbb{Z})$ is a rational function of $j(z)$.

Proof. Let $R = \{z \in \mathbb{H} : |\operatorname{Re}(z)| \leq 1/2, |\operatorname{Im}(z)| \geq 1/2\}$. $f(z)$ has only finitely many poles in R. Suppose there is a pole of order k at $\tau \in R$.
If $j'(\tau) \neq 0$, then $(j(z) - j(\tau))^k f(z)$ is holomorphic at τ.
If $j'(\tau) = 0$, then either $\tau = i$ or $\tau = e^{2\pi i/3}$. Suppose $\tau = i$. In a neighborhood of i, we have

$$f(z) = \frac{g(z)}{(z-i)^k},$$

where $g(z)$ is holomorphic and $g(i) \neq 0$. But $f(z)$ is invariant under $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, so

$$f(z) = f(-1/z) = \frac{g(-1/z)}{(-1/z - i)^k}.$$

We conclude that

$$g(-1/z) = \frac{g(z)}{(iz)^k},$$

which evaluated at i gives $g(i) = (-1)^k g(i)$. Since $g(i) \neq 0$, we must have that k is even. But $j(z) - 1728$ has a zero of order 2 at i, so $(j(z) - 1728)^{k/2} f(z)$ is holomorphic at i.
If $\tau = e^{2\pi i/3}$, one uses invariance with respect to $\begin{pmatrix} 1 & 1 \\ -1 & 0 \end{pmatrix}$ to show that k is a multiple of 3, then

$j(z)^{k/3}f(z)$ is holomorphic at τ.
So there exists a polynomial $Q(X)$ such that $Q(j(z))f(z)$ is holomorphic on R. But R contains a complete set of representatives for the action of $SL_2(\mathbb{Z})$, so by invariance $Q(j(z))f(z)$ is a holomorphic modular function for $SL_2(\mathbb{Z})$, i.e. a polynomial of $j(z)$. $\square$

Let $\mu = [SL_2(\mathbb{Z}) : \Gamma_0(N)]$, and write $\Gamma_0(N)\gamma_i$, $i = 1, \dots, \mu$ for the cosets of $\Gamma_0(N)$ in $SL_2(\mathbb{Z})$ ($\gamma_1 = 1$). Define

$$C(N) = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} : ad = N, a > 0, 0 \leq b < d, \gcd(a, b, d) = 1 \right\}.$$

The element $\sigma_0 = \begin{pmatrix} N & 0 \\ 0 & 1 \end{pmatrix}$ is distinguished by the fact that

$$\Gamma_0(N) = (\sigma_0^{-1} SL_2(\mathbb{Z}) \sigma_0) \cap SL_2(\mathbb{Z}).$$

There is a one-to-one correspondence between elements of $C(N)$ and cosets of $\Gamma_0(N)$, given by

$$\sigma \mapsto (\sigma_0^{-1} SL_2(\mathbb{Z}) \sigma) \cap SL_2(\mathbb{Z}).$$

Let $\gamma \in SL_2(\mathbb{Z})$ and choose $\sigma \in C(N)$ such that γ lies in the right coset corresponding to σ; there exists $\gamma' \in SL_2(\mathbb{Z})$ such that $\sigma_0 \gamma = \gamma' \sigma$. Therefore

$$j_N(\gamma z) = j(\sigma_0 \gamma z) = j(\gamma' \sigma z) = j(\sigma z). \tag{1}$$

Consider the following polynomial in X:

$$\Phi_N(X, z) = \prod_{i=1}^{\mu} (X - j_N(\gamma_i z)).$$

The coefficients of Φ_N are symmetric polynomials in the $j_N(\gamma_i z)$, so they are holomorphic on $\mathbb{H}$ and meromorphic at the cusps. Since the action of $SL_2(\mathbb{Z})$ simply permutes the terms in the product, they are also $SL_2(\mathbb{Z})$-invariant. Hence the coefficients of $\Phi_N(X, z)$ are holomorphic modular functions, i.e. polynomials of $j(z)$. So there exists a polynomial $\Phi_N(X, Y) \in \mathbb{C}[X, Y]$ such that

$$\Phi_N(X, j(z)) = \prod_{i=1}^{\mu} (X - j_N(\gamma_i z)). \tag{2}$$

By (1), we can write

$$\Phi_N(X, j(z)) = \prod_{\sigma \in C(N)} (X - j(\sigma z)).$$

But $j_N(z) = j(Nz) = j(\sigma_0 z)$, therefore

$$\Phi_N(j_N(z), j(z)) = 0.$$

$\Phi_N(X, Y) = 0$ is called the modular equation. $\Phi_N(X, Y)$ is irreducible with respect to X (hence it is the minimal polynomial of $j_N(z)$ over $\mathbb{C}(j(z))$).

Theorem 4. Every modular function for $\Gamma_0(N)$ is a rational function of $j(z)$ and $j_N(z)$.

Proof. Let $f(z)$ be a modular function for $\Gamma_0(N)$. Consider the polynomial in X

$$G(X, z) = \Phi_N(X, j(z)) \sum_{i=1}^{\mu} \frac{f(\gamma_i z)}{X - j_N(\gamma_i z)} = \sum_{i=1}^{\mu} f(\gamma_i z) \prod_{k \neq i} (X - j_N(\gamma_k z)).$$

Using an argument similar to the given above for Φ_N, one shows that the coefficients of G are modular functions for $SL_2(\mathbb{Z})$, hence rational functions of $j(z)$. That is, $G(X, j(z)) \in \mathbb{C}(j(z))[X]$. We differentiate (2)

$$\frac{\partial \Phi_N}{\partial X}(j_N(z), j(z)) = \prod_{i \neq 1} (j_N(z) - j_N(\gamma_i z)),$$

and get

$$G(j_N(z), j(z)) = f(z) \frac{\partial \Phi_N}{\partial X}(j_N(z), j(z)).$$

But $\Phi_N(X, j(z))$ is irreducible so the first derivative is nonzero. Finally

$$f(z) = \frac{G(j_N(z), j(z))}{(\partial/\partial X)\Phi_N(j_N(z), j(z))}.$$

□

3. Canonical model over $\mathbb{Q}$ and modular parametrization

We need some more information concerning the polynomial Φ_N. First note that invariance with respect to $\Gamma_0(N)$ gives

$$\Phi_N(j(z), j_N(z)) = \Phi_N(j(z), j(Nz)) = \Phi(j(-1/z), j(-1/Nz)) = \Phi_N(j_N(-1/Nz), j(-1/Nz)) = 0.$$

It is in fact true (and easy to show, see [Mil] p. 84) that $\Phi_N(X, Y)$ is symmetric.

Theorem 5. $\Phi_N(X, Y) \in \mathbb{Z}[X, Y]$.

Proof. Let $\zeta = e^{2\pi i/N}$. Since

$$j(z) = \frac{1}{q} + \sum_{n=0}^{\infty} c_n q^n,$$

where $c_n \in \mathbb{Z}$, and $j_N(\gamma_i z) = j(\sigma z) = j((az+b)/d)$ for some $\sigma = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in C(N)$, we get that $j_N(\gamma_i z)$ has a Fourier expansion in powers of $q^{1/N}$ whose coefficients are in $\mathbb{Z}[\zeta]$. Therefore any symmetric polynomial in the $j_N(\gamma_i z)$ has such an expansion with coefficients that are algebraic integers. We know that such polynomial lies in $\mathbb{C}[j(z)]$, and we claim that it has algebraic integer coefficients.
Suppose $S = \sum c_n j(z)^n \in \mathbb{C}[j(z)]$ has some coefficients that are not algebraic integers. Let c_k be the coefficient of this type, with largest index. Then the coefficient of q^{-k} in the q-expansion of S is not an algebraic integer, and hence S cannot be a symmetric polynomial in the $j_N(\gamma_i z)$.
So we know that $\Phi_N(X, Y) = \sum c_{m,n} X^m Y^n$ with algebraic integer $c_{m,n}$'s. We substitute the q-expansion of j into the modular equation $\Phi_N(j_N(z), j(z)) = 0$ and equate coefficients of powers of q. This gives a set of linear equations in the $c_{m,n}$ with rational coefficients. The $c_{m,n}$ are uniquely determined by this system, because there is only one monic minimal equation for $j_N(z)$ over $\mathbb{C}(j(z))$. The system has a solution in $\mathbb{C}$, and this solution is unique so it must lie in $\mathbb{Q}$. But we already know that $c_{m,n}$ are algebraic integers, so they must be in $\mathbb{Z}$. □

Note that except for the first few values of N, the polynomial Φ_N is not known explicitly (for $N = 11$, it has degree 21 and coefficients up to 10^{60}).

Since $X_0(N)$ is a compact Riemann surface, there is a unique structure of a nonsingular projective curve on $X_0(N)$ which is compatible with the conformal structure. We write $X_0(N)_{\mathbb{C}}$ for $X_0(N)$ viewed as an algebraic curve over $\mathbb{C}$. This is the unique nonsingular projective curve over $\mathbb{C}$ having $\mathbb{C}(j(z), j_N(z))$ as its field of rational functions.

Let C be the curve over $\mathbb{Q}$ defined by the modular equation $\Phi_N(X, Y) = 0$. C is singular, so we remove the singular points and embed the result into a nonsingular projective curve C'. The coordinate functions x and y generate the field of rational functions on C' and satisfy the relation $\Phi_N(x, y) = 0$. If $C'_{\mathbb{C}}$ is the curve defined by C' over $\mathbb{C}$, there is a unique isomorphism $C'_{\mathbb{C}} \to X_0(N)_{\mathbb{C}}$ such that the rational functions x and y correspond to $j_N(z)$ and $j(z)$. We identify the two curves via this isomorphism and regard C' as a model of $X_0(N)_{\mathbb{C}}$ over $\mathbb{Q}$. This is called the canonical model of $X_0(N)$ over $\mathbb{Q}$ and is denoted $X_0(N)_{\mathbb{Q}}$.

Theorem 6. For any elliptic curve E over $\mathbb{Q}$, there exists a positive integer N and a surjective morphism $\varphi : X_0(N) \to E$ defined over $\mathbb{Q}$.

We refer to the map φ as the modular parametrization of E.

4. Moduli interpretation of modular curves

Let k be a field. A moduli problem over k is a contravariant functor $\mathcal{F}$ from the category of varieties over k to the category of sets. Usually, $\mathcal{F}(V)$ is the set of isomorphism classes of certain objects over V.

A solution to the moduli problem $\mathcal{F}$ is a pair (V, α), where V is a variety over k and $\alpha : \mathcal{F}(k) \to V(k)$ is a bijection satisfying the following conditions:

1. Let T be a variety over k and $f \in \mathcal{F}(T)$. Any $t \in T(k)$ corresponds to a map $\operatorname{Maxspec}(k) \to V$, so f defines an element $f_t \in T(k)$. We have a map $t \mapsto \alpha(f_t)$ from $T(k)$ to $V(k)$ and this map is required to be a morphism.
2. Let Z be a variety over k and $\beta : \mathcal{F}(k) \to Z(k)$ be a map satisfying condition 1. Then $\beta \circ \alpha^{-1} : V(k) \to Z(k)$ is a morphism.

The point of the definition is that if a solution exists, we want it to give an identification $V(k) = \mathcal{F}(k)$ and to be unique.

Let V be a variety over $\mathbb{C}$. An elliptic curve over V is a morphism $E \to V$, where E is the subvariety of $V \times \mathbb{P}^1$ defined by a nonsingular Weierstrass equation with the a_i regular functions on V. We define $\mathcal{E}_{0,N}(V)$ to be the set of isomorphism classes of pairs (E, G) where E is an elliptic curve over V and G is a cyclic subgroup of E of order N.

Theorem 7. Let k be a field, N an integer not divisible by the characteristic of k. Then the moduli problem $\mathcal{E}_{0,N}$ has a solution (M, α) over k. When $k = \mathbb{Q}$, M is canonically isomorphic to $Y_0(N)_{\mathbb{Q}}$. The map

$$\mathcal{E}_{0,N}(k) \to M(k) = Y_0(N)_{\mathbb{Q}}(k)$$

is given by $(E, G) \mapsto (j(E), j(E/G))$.

Proof. See [Mil], pp. 94–95. □

5. Heegner points

Let E be an elliptic curve (without CM) over $\mathbb{Q}$, and fix a modular parametrization $\varphi : X_0(N) \to E$ which maps ∞ to 0. Let $K = \mathbb{Q}(\sqrt{-D})$ be an imaginary quadratic field of discriminant $-D$ ($D \neq 3, 4$) and where all prime factors of N are split, $(N) = \mathcal{N} \cdot \overline{\mathcal{N}}$. Let $\mathcal{O}$ be the ring of integers of K.

Fix an integer $n \geq 1$, prime to N and let $\mathcal{O}_n = \mathbb{Z} + n\mathcal{O}$ be the order of conductor n in $\mathcal{O}$. Let ω be a generator of $\mathcal{O}_n$ over $\mathbb{Z}$. The ideal $\mathcal{N}_n = \mathcal{N} \cap \mathcal{O}_n$ is an invertible $\mathcal{O}_n$-module with $\mathcal{O}_n/\mathcal{N}_n \cong \mathbb{Z}/N\mathbb{Z}$, so there exists an integer k, $0 \leq k \leq N-1$, such that $\omega - k \in \mathcal{N}_n$. $\{1, \omega - k\}$ generate $\mathcal{O}_n$ as a $\mathbb{Z}$-module, while $\{N, \omega - k\}$ generate $\mathcal{N}_n$ as a $\mathbb{Z}$-module. Then $\mathbb{C}/\mathcal{O}_n$ is an elliptic curve with CM by $\mathcal{O}_n$ and $G = \mathcal{N}_n^{-1}/\mathcal{O}_n$ is a cyclic subgroup of order N such that the quotient $\mathbb{C}/\mathcal{N}_n^{-1}$ also has CM by $\mathcal{O}_n$; so the point

$$x_n = \left(j\left(\frac{1}{\omega - k}\right), j_N\left(\frac{1}{\omega - k}\right) \right) \in X_0(N)$$

has coordinates lying in the ring class field K_n of $\mathcal{O}$ (without loss of generality, $(\omega - k)^{-1} \in \mathbb{H}$).

The x_n are called Heegner points. They yield points on the original elliptic curve E as follows. Let $y_n = \varphi(x_n)$. Since φ is defined over $\mathbb{Q}$ and x_n is defined over K_n, we have $y_n \in E(K_n)$. Put

$$y_{n,K} = \operatorname{Tr}_K^{K_n}(y_n) \in E(K).$$

Gross and Zagier proved that $y_{1,K}$ has infinite order if and only if the analytic rank of E/K is 1. Therefore, Heegner points can be used to construct non-torsion K-rational points on elliptic curves of rank 1. For details on how to implement the construction, see [Elk94]. The reason why they are of interest to us is that their images $\{y_n\}$ form an Euler system which can be used to bound the Selmer group of E.

References

[Cox89] David A. Cox. *Primes of the form $x^2 + ny^2$*. Wiley Interscience, 1989.

[Elk94] N. Elkies. Heegner point computations. In L. Adleman and M.-D. Huang, editors, *Algorithmic number theory*, number 877 in Lecture notes in computer science, pages 122–133. Springer, 1994.

[Gro91] Benedict Gross. Kolyvagin's work on modular elliptic curves. In J. Coates and M. Taylor, editors, *L-functions and arithmetic*, number 153 in LMS lecture note series, pages 235–256. Cambridge University Press, 1991.

[Mil] J. Milne. Modular functions and modular forms. Course notes available at www.jmilne.org.

[Roh97] David E. Rohrlich. Modular curves, Hecke correspondences, and L-functions. In G. Cornell, J. Silverman, and G. Stevens, editors, *Modular forms and Fermat's last theorem*, pages 41–100. Springer, 1997.