MODULAR CURVES AND HAUPTMODULN

ALEXANDER MIGRDITCHIAN

Abstract. By parameterizing their isomorphism classes, modular curves – closely related to modular forms – provide additional structure to elliptic curves. In this paper, we introduce the topological and algebraic properties of modular curves (and their compactifications) and subsequently proceed to introduce the modular j-function and Hauptmoduln. In doing so, we provide an exposition of the notion of modularity, which we use to discuss further interesting group-theoretic properties of modular curves.

1. Preliminary Definitions

We assume basic knowledge of group theory (in particular, the modular group $\mathrm{SL}_2(\mathbb{Z})$, sporadic groups, and normalizers) and familiarity with elliptic curves. As such, we begin by presenting the definition of modular curves:

Definition 1.1. A *modular curve* is defined to be a quotient of the upper-half plane $\mathbb{H}$ by the action of a congruence subgroup $\Gamma(N)$ of $\mathrm{SL}_2(\mathbb{Z})$ with finite index (that is, $\mathbb{H}\backslash\Gamma(N)$).

We have that two points are the same if they can be mapped to each other by Γ. Now, we require the definition of congruence subgroups before we can extract further results with respect to modular curves:

Definition 1.2. The *principal congruence subgroup* of level N is defined to be

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

Note that $\Gamma(1) = \mathrm{SL}_2(\mathbb{Z})$. There are congruence subgroups other than the principle congruence subgroup, the most important of which are $\Gamma_0(N)$ and $\Gamma_1(N)$ as they satisfy the relation $\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N) \subset \mathrm{SL}_2(\mathbb{Z})$. We define these congruence subgroups as follows:

Definition 1.3. A *congruence subgroup* Γ of $\mathrm{SL}_2(\mathbb{Z})$ satisfies $\Gamma(N) \subset \Gamma$ for a positive integer N. In particular, Γ is a congruence subgroup of level N.

Definition 1.4. The congruence subgroup $\Gamma_0(N)$ is defined as

$$\Gamma_0(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : c \equiv 0 \pmod N \right\}.$$

Definition 1.5. The congruence subgroup $\Gamma_1(N)$ is defined as

$$\Gamma_1(N) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}) : c \equiv 0 \pmod N, a \equiv d \equiv 1 \pmod N \right\}.$$

Date: June 6, 2021.

We may now begin our exposition of modular curves. We have that the modular curve $Y(\Gamma)$ is the quotient space of orbits under the action of Γ, namely

$$Y(\Gamma) = \Gamma\backslash\mathbb{H} = \{\Gamma\tau : \tau \in \mathbb{H}\}.$$

This quotient intuitively makes sense because of the compactification of Γ, which we will explain later. Indeed, we have that the modular curves for the aforementioned congruence subgroups are $Y_0(N) = \Gamma_0(N)\backslash\mathbb{H}$, $Y_1(N) = \Gamma_1(N)\backslash\mathbb{H}$, and $Y(N) = \Gamma(N)\backslash\mathbb{H}$. This set of orbits has an interesting topological interpretation: that is, the notion that a Riemann surface can be compactified. We denote a compactified Riemann surface $Y(\Gamma)$ by $X(\Gamma)$.

2. Topological Notions: Modular Curves as Riemann Surfaces

Firstly, we consider the surjection $\pi : \mathbb{H} \mapsto Y(\Gamma)$ with $\pi(\tau) = \Gamma\tau$, because this gives $Y(\Gamma)$ a quotient topology: that is, a subset of $Y(\Gamma)$ is open if its inverse image $\pi^{-1}[Y(\Gamma)]$ under π in $\mathbb{H}$ is also open. In particular, this is the weakest topology (i.e., the topology with the minimal number of open sets) such that $\pi : \mathbb{H} \mapsto Y(\Gamma)$ is continuous.

Thus, it is evident that π is an open mapping, from which we present the following proposition, which in turn allows for us to conclude that the quotient $Y(\Gamma)$ is connected for continuous π (and since $\mathbb{H}$ is connected):

Proposition 2.1. *$\pi(U_1) \cap \pi(U_2) = \emptyset$ in $Y(\Gamma)$ if and only if $\Gamma(U_1) \cap U_2 = \emptyset$ in $\mathbb{H}$ for $U_1, U_2 \subset \mathbb{C}$.*

Perhaps the most interesting property of $Y(\Gamma)$ is that it is Hausdorff, from which we have the following definition:

Definition 2.2. A Hausdorff (T_2) space is a topological space in which any two distinct points have neighborhoods that are disjoint from each other.

We show this result by taking sufficiently small neighborhoods of $\mathbb{H}$ under charts on $Y(\Gamma)$ for which every $\mathrm{SL}_2(\mathbb{Z})$ transformation on $\mathbb{H}$ is properly discontinuous. To this end, we present the following lemma:

Lemma 2.3. *Take $\tau_1, \tau_2 \in \mathbb{H}$ and $U_1, U_2 \subset \mathbb{C}$ that are neighborhoods of τ_1 and τ_2, respectively, in $\mathbb{H}$. Then if we take $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, we have that*

$$\gamma(U_1) \cap U_2 \neq \emptyset \implies \gamma(\tau_1) = \tau_2.$$

In turn, this allows for us to prove our result by considering such neighborhoods and points $\pi(\tau_1)$ and $\pi(\tau_2)$ in $Y(\Gamma)$.

Theorem 2.4. *The modular curve $Y(\Gamma)$ is a Hausdorff space for any congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$.*

Proof. Take the aforementioned distinct points in $Y(\Gamma)$ as well as their neighborhoods. For any choice of $\gamma \in \Gamma$, we have that $\gamma(\tau_1) \neq \tau_2$, and from Lemma 2.3, we have that

$$\gamma(U_1) \cap U_2 = \emptyset$$

in $\mathbb{H}$. In particular, from Proposition 2.1, it is clear that $\pi(U_1)$ and $\pi(U_2)$ are disjoint supersets of $\pi(\tau_1)$ and $\pi(\tau_2)$ under $Y(\Gamma)$. Since π is also an open mapping under $\mathbb{H}$, we conclude the desired result. ■

Since we have shown that $Y(\Gamma)$ is Hausdorff, we seek to map charts onto the modular curve $Y(\Gamma)$, which we in turn do by introducing isotropy subgroups and elliptic points.

Definition 2.5. The *isotropy subgroup* of τ fixes τ on some congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$:

$$\Gamma_\tau = \{\gamma \in \Gamma : \gamma(\tau) = \tau\}.$$

Definition 2.6. An *elliptic point* $\tau \in \mathbb{H}$ of Γ satisfies a relation with respect to the containment of matrix groups, namely $\{\pm I\} \subset \{\pm I\}\Gamma_\tau$ for the identity matrix I. We must have that Γ_τ is nontrivial as a group of transformations. In particular, we also have that the point $\pi(\tau) \in Y(\Gamma)$ is elliptic.

One important result is that $Y(\Gamma)$ only has finitely many elliptic points; we show this by taking a congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$ for which each elliptic point $\tau \in \mathbb{H}$ of Γ has an isotropy subgroup Γ_τ that is finite and cyclic. For the sake of brevity, we will not provide an exposition of this result for all N (we encourage the reader to see [DS05]), but we will introduce the simple case of $Y(1) = \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$. If we define the *fundamental domain of* $\mathrm{SL}_2(\mathbb{Z})$ (see Figure 1) as

$$\mathcal{D} := \{\tau \in \mathbb{H} : |\Re(\tau)| \leq 1/2, |\tau| \geq 1\},$$

we have the following result:

Theorem 2.7. *For the natural projection π with $\pi(\tau) = \mathrm{SL}_2(\mathbb{Z})\tau$, we have that the map $\pi : \mathcal{D} \mapsto Y(1)$ is surjective.*

Proof. The result follows from repeated actions of the matrices

$$\begin{pmatrix} 1 & \pm 1 \\ 0 & 1 \end{pmatrix} \text{ and } \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

to $\tau \in \mathbb{H}$, which in turn map $\tau \mapsto \tau \pm 1$ and $\tau \mapsto -1/\tau$, respectively. Repeatedly applying these matrices to $\tau \notin \mathcal{D}$ will eventually terminate with τ being contained within the region $\mathcal{D}$ as defined above. ■

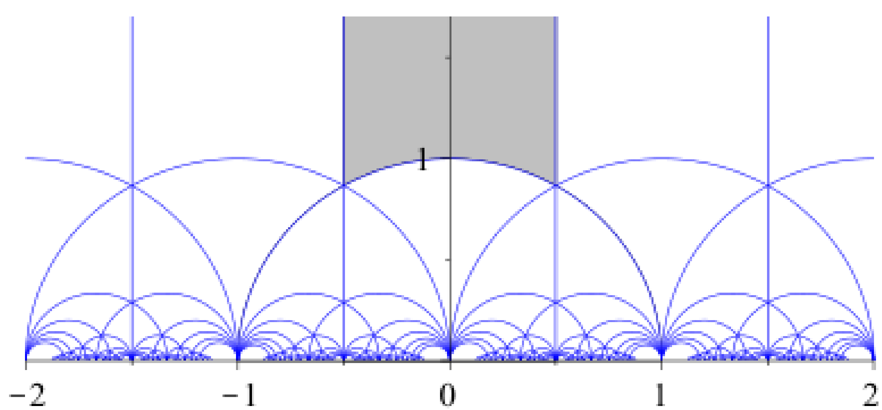

Figure 1. The fundamental domain of the modular group $\mathrm{SL}_2(\mathbb{Z})$, sometimes denoted by R_Γ.

3. Cusps

The basic notion of cusps is as follows: for a congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$, we have that the cusps of Γ are the Γ-equivalence classes of $\mathbb{Q} \cup \{\infty\}$. We can create a compact Riemann surface $X(\Gamma)$ by adjoining cusps and charts to the modular curve $Y(\Gamma)$: in particular, we also have that the compactified $X(\Gamma)$ is a modular curve. Motivated by this, we continue our exposition of compactification of the modular curve $Y(\Gamma) = \Gamma\backslash\mathbb{H}$.

Definition 3.1. The *extended complex upper-half plane* is defined by $\mathbb{H}^* := \mathbb{H} \cup \mathbb{Q} \cup \{\infty\}$.

We may now consider the extended quotient

$$X(\Gamma) = \Gamma\backslash\mathbb{H}^* = Y(\Gamma) \cup \Gamma\backslash(\mathbb{Q} \cup \{\infty\}),$$

from which we present the formal definition of cusps:

Definition 3.2. The points Γs in $\Gamma\backslash(\mathbb{Q} \cup \{\infty\})$ are the *cusps* of the modular curve $X(\Gamma)$.

Note that the action of Γ on $\mathbb{Q} \cup \{\infty\}$ decomposes it into orbits, which are precisely the cusps of $X(\Gamma)$. The additional case of a transitive action of Γ on $\mathbb{Q} \cup \{\infty\}$ is termed the *Alexandroff compactification* of $\Gamma\backslash\mathbb{H}$ and is beyond the scope of this paper, but we encourage the reader to read up further on the topic.

We correspondingly define the modular curves $X_0(N)$, $X_1(N)$, and $X(N)$ for the congruence subgroups $\Gamma_0(N)$, $\Gamma_1(N)$, and $\Gamma(N)$. Now, we will present two important results pertaining to the number of cusps contained by a modular curve and another important property of $X(\Gamma)$. The latter result requires an understanding of topology that is beyond the scope of this self-contained paper, so we provide brief motivation behind the topology on $X(\Gamma)$.

Proposition 3.3. *Take a congruence subgroup Γ of $\mathrm{SL}_2(\mathbb{Z})$. Then, the modular curve $X(\Gamma)$ has finitely many cusps, and in particular, the modular curve $X(1)$ is equivalent to $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}^*$.*

Take a real $M > 0$ and correspondingly define the neighborhood

$$\mathcal{N}_M := \{\tau \in \mathbb{H} : \Im(\tau) > M\}.$$

By adjoining open sets in $\mathbb{H}$ to $\mathbb{H}^*$, we generate a basis for the neighborhoods of cusps, namely sets that take the form

$$\alpha(\mathcal{N}_M \cup \{\infty\}) : \alpha \in \mathrm{SL}_2(\mathbb{Z}), M > 0.$$

In doing such, we are able to define a topology and $\mathbb{H}^*$. Indeed, we have that a basis

$$\mathcal{B} := \{\tau \in \mathbb{H} : \Im(\tau) > M\} \subset \mathbb{H}$$

with $M > 0$ is a topology on $\mathbb{H}^*$, motivating the following result:

Proposition 3.4. *The modular curve $X(\Gamma)$ is compact, connected, and Hausdorff.*

Interestingly, the number of cusps of $X_N(\Gamma)$ is closely related to the totient function φ. For example, we have the following result that quantifies the number of cusps on the modular curve $X_1(\Gamma)$:

Proposition 3.5. *The number of cusps on $X_1(\Gamma)$ is*

$$\frac{1}{2}\sum \varphi(d)\varphi(N/d) = \frac{N}{2}\prod_{p|N}\left(1 - 1/p^2 + \nu_p(N)(1-1/p)^2\right),$$

where $\nu_p(N) := \sum_{k\geq 1}\lfloor N/p^k \rfloor$ is the p-adic valuation of N.

We refer the reader to [DI95] for an exposition of the proof.

Since we have introduced the basic Riemann surface theory of the compactified modular curve $X(\Gamma)$, we could continue further by introducing the dimension formulas, the genus of $X(\Gamma)$, and other results, such as the Riemann-Roch Theorem. However, from this point in the paper, we will focus less on the topological properties of modular curves and more so on their algebraic properties.

Our discussion of cusps will prove useful later as we introduce cusp forms, from which we will introduce the j-function through the Eisenstein series, which is closely related to cusp forms.

4. Modular Curves as Algebraic Curves

We can describe modular curves as algebraic curves by taking advantage of the fact that curves are characterized by fields. Moreover, we can describe mappings between such curves by considering field extensions. As such, we make a few preliminary definitions before introducing the notions of a coordinate ring and a polynomial function on a function field, which will prove important in characterizing modular curves.

Consider a field $\mathbf{k}$ and positive integers m, n. Now, take polynomials $\varphi_1, \ldots, \varphi_m \in \mathbf{k}[x_1, \ldots, x_n]$ and denote by I the ideal generated by $\varphi_1, \ldots, \varphi_m$ in $\mathbf{k}[x_1, \ldots, x_n]$ (namely, the ring of polynomials over the algebraic closure of the field). That is, we have

$$I = \langle \varphi_1, \ldots, \varphi_n \rangle \subset \mathbf{k}[x_1, \ldots, x_n].$$

Moreover, if we denote by C the set of solutions to the polynomials in the ideal, we have that

$$C = \{P \in \overline{\mathbf{k}}^n : \varphi(P) = 0 \ \forall\ \varphi \in I\}.$$

Now, we may present our definitions:

Definition 4.1. The *coordinate ring* of a set C over a field $\mathbf{k}$ is defined to be the integral domain

$$\mathcal{O}(C) = \overline{\mathbf{k}}[C] = \overline{\mathbf{k}}[x_1, \ldots, x_n]/I.$$

In particular, a $P \in \mathcal{O}(C)$ is a *polynomial function* on C.

We may further define the notion of function fields:

Definition 4.2. The *function field* of a set C over the algebraic closure of a field $\mathbf{k}$ is the quotient field of the coordinate ring. That is,

$$\overline{\mathbf{k}}(C) := \{f/g : f, g \in \overline{\mathbf{k}}[C], g \neq 0\}.$$

Perhaps the most interesting result of this is the relationship between the function field $\mathbb{C}[X(1)]$ (recall that $X(1) = \mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}^*$) and the j-function (which we will define in the subsequent section) by direct equivalence: indeed, we can generate $\mathbb{C}[X(1)]$ with the field $\mathbb{C}[X(j)]$. To do so, however, we must first introduce the motivation behind the j-function through cusps and the Eisenstein series.

5. More on Cusps: Cusp Forms and Modularity

Recall that the Weirstrauss $\wp$ function is defined as follows for a choice of lattice Λ and $\omega \in \mathbb{C}$:

$$\wp_\Lambda(z) = \frac{1}{z} + \sum_{\omega \in \Lambda} \left(\frac{1}{(z-\omega)^2} - \frac{1}{\omega^2} \right).$$

We now introduce the Eisenstein series, which is the two-dimensional analog of the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$ as follows:

Definition 5.1. For some $\tau \in \mathbb{H}$, the *Eisenstein series* of weight k is defined as

$$G_k(\tau) = \sum_{(c,d) \in \mathbb{Z}^2 \setminus \{(0,0)\}} \frac{1}{(c\tau + d)^k}.$$

Now define the functions $g_2(\tau) := 60G_4(\tau)$ and $g_3(\tau) = 140G_6(\tau)$, as well as the discriminant function $\Delta : \mathbb{H} \mapsto \mathbb{C}$ by $\Delta(\tau) := g_2(\tau)^3 - 27g_3(\tau)^2$. We may now introduce the modular j-function, also from $\mathbb{H} \to \mathbb{C}$ and which is holomorphic on $\mathbb{H}$:

Definition 5.2. The j-function is defined by

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

One immediate property of the j-function is that for a choice of $\tau \in \mathbb{H}$ and $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, we have that $j(\tau) = j(\gamma(\tau))$, i.e. that the j-function is invariant under the action of $\mathrm{SL}_2(\mathbb{Z})$. This follows from the numerator and denominator of $j(\tau)$ (namely, $g_2(\tau)$ and $\Delta(\tau)$) being Eisenstein series of the same weight. Additionally, note that

$$\lim_{\Im(\tau) \to \infty} j(\tau) = \infty.$$

Brief reasoning for this result can be found in [Sch10]. It follows that $j(\tau)$ is a modular function on $\mathrm{SL}_2(\mathbb{Z})$, from which it can be shown that $j : \mathbb{H} \mapsto \mathbb{C}$ is surjective. More generally, the aforementioned functions $g_2(\tau)$ and $g_3(\tau)$ can be characterized as cusp forms of weight $k = 2$ and 3, respectively, the set of which is denoted by $\mathcal{S}_k(\mathrm{SL}_2(\mathbb{Z}))$:

Definition 5.3. A *cusp form* of weight k is a modular curve with a Fourier expansion that takes the following form:

$$f(\tau) = \sum_{n=1}^{\infty} a_n e^{2\pi i \tau n}$$

with $a_0 = 0$.

6. Properties of the j-Function

If we take $q = e^{2\pi i \tau}$, we have the following *q-expansion* for $j(\tau)$:

$$j(\tau) = 1/q + 744 + 196884q + 21493760q^2 + \cdots = 1/q + \sum_{n=0}^{\infty} a_n q^n,$$

which turns out to be closely related to the Fischer-Greiss Monster group: the largest sporadic simple group, having order

$$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71 \approx 8 \times 10^{53}.$$

Motivated by this, we will introduce and use two results to provide an exposition of the interesting group theoretic properties of the j-function, namely:

Theorem 6.1. *The j-function is a bijection from* $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H} \mapsto \mathbb{C}$*.*

The result follows from considering two homothetic lattices Λ and $\Lambda' := \gamma\Lambda$ in $\mathbb{H}$ for some $\gamma \in \mathrm{SL}_2(\mathbb{Z})$, from which it can be shown that $j(\Lambda) = j(\Lambda')$ if and only if this is the case; a complete proof can be found in [Sch10].

Theorem 6.2. *Additionally, every modular function is a rational function of $j(\tau)$.*

While the proof of the latter result will be eliminated for brevity, it follows from considering the Laurent series of the q-expansion of $j(\tau)$, which takes the form $f(\tau) = \sum_{n=-m}^{\infty} a_n q^n$ since f is modular and meromorphic at the cusp $i\infty$. From this, we have that every holomorphic function for $\mathrm{SL}_2(\mathbb{Z})$ is a polynomial function in $j(\tau)$, and the desired result follows.

The latter result is especially powerful because of how it allows for us to view $j(\tau)$ as the generator of all meromorphic modular functions, termed *Hauptmodul*. Since formally introducing Hauptmodul will require us to provide an exposition of further topological definitions (such as the definition of a topological genus), we will now begin to introduce the connections between the j-function and the Monster group.

Firstly, the notion of *genus zero* refers to function fields with one transcendental function as a generator. This is particularly relevant in the case of *modular function fields*, which are function fields of modular curves. Indeed, as an example, we have the following result (here, the projective special linear group $\mathrm{PSL}_2(\mathbb{Z})$ is the analogous action of $\mathrm{SL}_2(\mathbb{Z})$ on the associated projective space):

Proposition 6.3. *The function field of the modular curve* $X(1) := \mathrm{PSL}_2(\mathbb{Z})\backslash\mathbb{H}^*$ *is genus zero. In particular, it is generated by the modular j-function.*

We may now formally define Hauptmodul:

Definition 6.4. A *Hauptmodul* is a principal modular function that generates a function field, unique up to a normalizable Möbius transformation.

The connections between modular curves, the j-function, and the Monster group become particularly evident as one considers modular curves of genus zero and the coefficients of the q-expansions of their Hauptmoduln. If we take a prime factor of the order of the Monster group

$$p \in \{2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 41, 47, 59, 71\}$$

we conclude the following result relating the normalizer of a modular curve in $\mathrm{SL}_2(\mathbb{R})$ with its genus:

Theorem 6.5. *The normalizer $\Gamma_0(p)^+$ of $\Gamma_0(p)$ in $\mathrm{SL}_2(\mathbb{Z})$ has genus 0 if and only if p is a prime factor of the order of the Fischer-Greiss Monster group.*

References

[DI95] Fred Diamond and John Im. *Modular forms and modular curves.* Canadian Mathematical Society, 1995. URL: `https://people.math.wisc.edu/~boston/Diamond-Im-Modular_forms_and_modular_curves.pdf`.

[DS05] Fred Diamond and Jerry Shurman. *A First Course in Modular Forms.* Graduate Texts in Mathematics. Springer New York, 2005.

[Fd17] Fundamental domain for the modular group. 2017. URL: https://en.wikipedia.org/wiki/Fundamental_domain.

[Ogg74] Andrew P. Ogg. *Automorphismes de courbes modulaires*. Seminaire Delange-Pisot-Poitou. Theorie des nombres, tome 16, no. 1 (1974–1975), exp. no. 7 (in French), 1974.

[Sch10] Asa Scherer. The j-function and the monster. 2010. URL: http://sites.science.oregonstate.edu/~swisherh/AsaScherer.pdf.

[Shi94] Goro Shimura. *Introduction to the arithmetic theory of automorphic functions*. Kanô Memorial Lectures. Publications of the Mathematical Society of Japan, Princeton University Press, 1994.

Euler Circle
Email address: alexmigrditchian@gmail.com