M. V. Hoeij, H. Smith *Res. Number Theory* (2021) 7:22
https://doi.org/10.1007/s40993-021-00243-3

Research in Number Theory

RESEARCH

A Divisor Formula and a Bound on the $\mathbb{Q}$-Gonality of the Modular Curve $X_1(N)$

Mark van Hoeij[1] and Hanson Smith[2*]

*Correspondence:
hanson.smith@uconn.edu
[2]Department of Mathematics, University of Connecticut, 341 Mansfield Road U1009, Storrs, CT 06269-1009, USA
Full list of author information is available at the end of the article

Abstract

We give a formula for divisors of modular units on $X_1(N)$ and use it to prove that the $\mathbb{Q}$-gonality of the modular curve $X_1(N)$ is bounded above by $[11N^2/840]$, where $[\cdot]$ denotes the nearest integer.

Keywords: Modular curves, Gonality, modular units, Siegel functions, Torsion points on elliptic curves

Mathematics Subject Classification: Primary 11G16, Secondary 14H52, 11G05, 14G35, 11F03

1 Introduction

The modular curve $X_1(N)$ parametrizes pairs $(E, \pm P)$ where E is an elliptic curve and P is a point of exact order N. As such it has been an object of interest for number theorists and arithmetic geometers. If K is a field, then *K-gonality* of $X_1(N)$ is the minimum degree of a non-constant function $X_1(N) \to \mathbb{P}^1$ defined over K.

Table 1 in [6] gives the currently-best upper bounds on the $\mathbb{Q}$-gonality of $X_1(N)$ for $N \leq 250$ and matching lower bounds for $N \leq 40$. Any non-constant function provides an upper bound for the gonality. The upper bounds in [6, Table 1] come from *modular units*. These are functions on $X_1(N)$ whose divisors are supported only on *cusps* (places on $X_1(N)$ where E degenerates). In this note we prove a formula for the degree of a certain modular unit F_7/F_8. Its degree is a particularly good gonality bound when N is prime. For all primes $N \leq 250$ except 31, 67, 101 it equals the currently the best upper bound from [6, Table 1].

Modular units are usually given in terms of Siegel functions [5, Theorem 3.4]. Conjecture 1 in [6], which was proved in [18], gave a basis $F_2, F_3, \ldots$ of modular units in algebraic form, which is useful for computer computations. In order to quickly find the degree of any modular unit, a formula for the divisor of $F_k : X_1(N) \to \mathbb{P}^1$ was given at [25].

A proof for this formula was not given; the resulting degrees listed in [6, Table 1] were verified by other means. The main result in this paper is a proof for this formula (Theorem 1 in Sect. 4), the "MinFormula." As an application, Sect. 5 gives the bound

$$\text{Gonality}_{\mathbb{Q}}(X_1(N)) \leq \deg\left(\frac{F_7}{F_8} : X_1(N) \to \mathbb{P}^1\right) \leq \left[\frac{11N^2}{840}\right] \quad \text{if } N > 8.$$

Here $[\cdot]$ indicates rounding to the nearest integer. The second $\leq$ is an equality when N is prime. The asymptotic growth $11N^2/840$ was already observed in [6, Section 2.1] and [19, p. 11] (combine the factors 11/35 and 1/24) though a proof was not given.

The explicit divisors given in Theorem 1 have other applications as well, such as computing Galois representations for modular curves [7], computing the action of diamond operators [6], computing cuspidal class numbers of modular curves [3,4], [12, Chapters 5 and 6], [23,27–30], computing optimized equations for $X_1(N)$ [1,20,21], and sporadic points on modular curves [2,8,13,14,17,19,24].

Section 2 reviews Puiseux expansions, elliptic curves, and division polynomials, and the relation to modular units is given in Sect. 3. To help the reader keep track of the notations we give a summary in Tables 1 and 2. Section 4 gives the main theorem. In Sect. 5 we obtain the gonality bound as an application of the main theorem. Streng [18] used Siegel functions to prove [6, Conjecture 1]. This work implies another proof for Theorem 1, see Appendix A for details. Orders of Siegel functions are typically expressed in terms of Bernoulli polynomials. We observe that such expressions sum to piecewise linear functions (Appendix A.3) when the corresponding product of Siegel functions is a modular unit.

The ideas for Theorem 1 are as follows: The defining polynomial q_k of $X_1(k)$ is a factor of the k^{th} division polynomial Q_k. A basis $F_2, F_3, \ldots$ of modular units of $X_1(N)$ is written in terms of these q_k. Theorem 1 is a formula for their divisors, obtained by computing valuations of q_k at cusps of $X_1(N)$ with $N \neq k$. In general, valuations can be determined from Puiseux expansions. In our case these expansions have distinct leading terms. Then only leading terms are needed, see Eq. (2). Leading terms correspond to asymptotic behavior, and cusps correspond to degenerate elliptic curves. When $\epsilon \to 0$, the elliptic curve $y^2 = x(x-\epsilon)(x-1)$ degenerates, and we compute the asymptotic behavior of its integrals in Sect. 2.2. Converting the resulting leading terms to the universal elliptic curve in Eq. (22) allows us to generalize Example 3 to Theorem 1.

2 Preliminaries

2.1 Places and Puiseux expansions

If $f \in \mathbb{Q}(s)[x]$ is irreducible over $\overline{\mathbb{Q}}$, then f defines an algebraic curve C whose function field $\mathbb{Q}(C)$ is $\mathbb{Q}(s)[x]/(f)$.

We give a brief summary of Puiseux expansions, see [15, Chapter II] for more. A *Puiseux expansion* of f at $s = 0$ is a root of f in the algebraic closure of $\mathbb{Q}((s))$. This is contained in the algebraic closure of $\mathbb{C}((s))$, which is $\bigcup_{e=1}^{\infty} \mathbb{C}\left(\left(s^{1/e}\right)\right)$. The natural valuation

$$v_s : \mathbb{C}\left(\left(s^{1/e}\right)\right) \to \frac{1}{e}\mathbb{Z} \bigcup \{\infty\}$$

sends a non-zero series to its lowest exponent in s and sends 0 to ∞.

For a Puiseux expansion $\mathbf{p}$, let $\mathbf{e_p}$ be the smallest positive integer e for which $\mathbf{p} \in \mathbb{C}\left(\left(s^{1/e}\right)\right)$. From the embedding

$$\phi_{\mathbf{p}} : \mathbb{Q}(C) \to \mathbb{Q}((s))[\mathbf{p}] \subset \mathbb{C}\left(\left(s^{1/\mathbf{e_p}}\right)\right), \quad \phi_{\mathbf{p}} : x \mapsto \mathbf{p}$$

we get a discrete valuation

$$v_{\mathbf{p}} : \mathbb{Q}(C) \to \mathbb{Z} \bigcup \{\infty\} \quad \text{given by} \quad v_{\mathbf{p}}(a) = \mathbf{e_p} \cdot v_s\left(\phi_{\mathbf{p}}(a)\right). \tag{1}$$

The factor $\mathbf{e_p}$ in (1) ensures that $\nu_{\mathbf{p}}(a)$ lands in $\mathbb{Z} \cup \{\infty\}$. Omitting this factor gives what we will call the *unweighted order* $v_s\left(\phi_{\mathbf{p}}(a)\right)$ of a, which is 1 at $a = s$ and $1/\mathbf{e_p}$ at a *local parameter*. The *residue field* $k_{\mathbf{p}}$ is defined as $\left\{a \in K_{\mathbf{p}} \mid \nu_{\mathbf{p}}(a) \geq 0\right\}$ modulo $\left\{a \in K_{\mathbf{p}} \mid \nu_{\mathbf{p}}(a) > 0\right\}$, where $K_{\mathbf{p}} = \mathbb{Q}((s))[\mathbf{p}]$.

A *place* on $C/\mathbb{Q}$ is a discrete valuation $\nu_P : \mathbb{Q}(C) \to \mathbb{Z} \bigcup \{\infty\}$. A place above $s = 0$ is a place with $\nu_P(s) > 0$. Puiseux expansions $\mathbf{p}$ and $\mathbf{p}_1$ are conjugate over $\mathbb{Q}((s))$ if and only if $\nu_{\mathbf{p}} = \nu_{\mathbf{p}_1}$, so *a place corresponds to a conjugacy class of Puiseux expansions.* A conjugacy class $\{\mathbf{p}, \ldots\}$ has $\mathbf{n_p} := \mathbf{e_p}\,\mathbf{f_p}$ elements, where $\mathbf{f_p} = [k_{\mathbf{p}} : \mathbb{Q}]$. A valuation $\nu_{\mathbf{p}} : \mathbb{Q}(C) \to \mathbb{Z} \bigcup \{\infty\}$ extends to $\mathbf{f_p}$ distinct valuations $\mathbb{C}(C) \to \mathbb{Z} \bigcup \{\infty\}$, so *one place on $C/\mathbb{Q}$ corresponds to $\mathbf{f_p}$ places on $C/\mathbb{C}$.*

Example 1 Let $\mathbf{p} = cs^{1/2} + \cdots$ where $c \neq 0$ and dots are terms of higher order. Then $v_s(\mathbf{p}) = 1/2$ so $\mathbf{p}^2/s = c^2 s^0 + \cdots$ has valuation 0 and hence $c^2 \in k_{\mathbf{p}}$. However, c need not be in $k_{\mathbf{p}}$. In that case, to avoid constants not in $k_{\mathbf{p}}$, we rewrite $cs^{1/2}$ as $(\alpha s)^{1/2}$ where $\alpha = c^2 \in k_{\mathbf{p}}$.

Definition 1 Let $l_s(\mathbf{p})$ denote the *dominant term* (the term with lowest exponent) of a nonzero Puiseux expansion. We write $\mathbf{p}_1 \sim \mathbf{p}_2$ if and only if $l_s(\mathbf{p}_1) = l_s(\mathbf{p}_2)$. In general, $v_s(\mathbf{p}_1 - \mathbf{p}_2) \geq \min(v_s(\mathbf{p}_1), v_s(\mathbf{p}_2))$ with equality if and only if $\mathbf{p}_1 \nsim \mathbf{p}_2$.

Let P be a place above $s = 0$ given by a Puiseux expansion $\mathbf{p} \in \mathbb{C}\left(\left(s^{1/\mathbf{e_p}}\right)\right)$ of f. Suppose we wish to compute $\nu_P(\overline{g})$ for some $g \in \mathbb{Q}(s)[x]$, where $\overline{g}$ is the image of g in $\mathbb{Q}(s)[x]/(f)$. Write $g = l\,(x - \mathbf{p}_1) \cdots (x - \mathbf{p}_n)$, where $l \in \mathbb{Q}(s)$ and the $\mathbf{p}_i \in \mathbb{C}((s^{1/\mathbf{e}_{\mathbf{p}_i}}))$ are the Puiseux expansions of g at $s = 0$. Then $g(\mathbf{p}) = l\,(\mathbf{p} - \mathbf{p}_1) \cdots (\mathbf{p} - \mathbf{p}_n)$ and $\nu_P(\overline{g}) = \mathbf{e_p} \cdot (v_s(l) + v_s(\mathbf{p} - \mathbf{p}_1) + \cdots + v(\mathbf{p} - \mathbf{p}_n))$. If $\mathbf{p} \nsim \mathbf{p}_i$ for each i, then:

$$\nu_P(\overline{g}) = \mathbf{e_p} \cdot \left(v_s(l) + \min\left\{v_s(\mathbf{p}), v_s(\mathbf{p}_1)\right\} + \cdots + \min\left\{v_s(\mathbf{p}), v_s(\mathbf{p}_n)\right\} \right). \tag{2}$$

Lemma 1 *With g and $\mathbf{p}_i$ as above, let $\mathbf{l}_i := l_s(\mathbf{p}_i)$. Suppose that $\mathbf{l}_1, \ldots, \mathbf{l}_n$ are distinct. Then $\mathbf{e}_{\mathbf{l}_i} = \mathbf{e}_{\mathbf{p}_i}$ and $k_{\mathbf{l}_i} = k_{\mathbf{p}_i}$.*

Proof Note $\mathbf{e}_{\mathbf{l}_i} \leq \mathbf{e}_{\mathbf{p}_i}$ and $k_{\mathbf{l}_i} \subseteq k_{\mathbf{p}_i}$ because the ramification index and residue field of $\mathbf{p}_i$ must be at least as large as those of its dominant term $\mathbf{l}_i$. If at least one of those is not an equality, then $\mathbf{n}_{\mathbf{p}_i} > \mathbf{n}_{\mathbf{l}_i}$. In this case, $\mathbf{p}_i$ has more conjugates over $\mathbb{Q}((s))$ than $\mathbf{l}_i$, so there must be at least two conjugates with the same dominant term. Those conjugates are among $\mathbf{p}_1, \ldots, \mathbf{p}_n$ since $g \in \mathbb{Q}(s)[x]$, which implies that $\mathbf{l}_1, \ldots, \mathbf{l}_n$ are not distinct.

2.2 Elliptic curves, analytic viewpoint

Let $0 < \epsilon \ll 1$ and consider the elliptic curve

$$E_\epsilon : y^2 = x(x - \epsilon)(x - 1) \tag{3}$$

so $y = \sqrt{x(x - \epsilon)(x - 1)}$. Let $E_\epsilon(\mathbb{C})$ denote the points on E defined over $\mathbb{C}$. This is an additive group, the identity $\mathcal{O}$ is the point at infinity. The period lattice is $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$

where [16, Chapter VI]

$$\omega_1 = 2\int_1^{\infty} \frac{\mathrm{d}x}{y} = 2\int_0^{\epsilon} \frac{\mathrm{d}x}{y} = 4K\left(\sqrt{\epsilon}\right) = 2\pi\left(1 + \frac{1}{4}\epsilon + \frac{9}{64}\epsilon^2 + \frac{25}{256}\epsilon^3 + \cdots\right) \tag{4}$$

and

$$\omega_2 = 2\int_{\epsilon}^{1} \frac{\mathrm{d}x}{y} = 2\int_{-\infty}^{0} \frac{\mathrm{d}x}{y} = \frac{4}{i}K\left(\sqrt{1-\epsilon}\right) = \frac{\omega_1}{\pi i}\ln\left(\frac{16}{\epsilon} - 8 - \frac{5}{4}\epsilon + \cdots\right). \tag{5}$$

Here K is the complete elliptic integral of the first kind [10, §19.2(ii)]

$$K(t) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1 - t^2\sin^2\theta}}.$$

In this section $\sim$ means that the ϵ-dominant terms are the same, similar to Definition 1. For example

$$\omega_1 \sim 2\pi \quad \text{and} \quad \omega_2 \sim \frac{2}{i}\ln\left(\frac{16}{\epsilon}\right). \tag{6}$$

The notation $\approx$ will be used for approximations in intermediate steps, to indicate that they are sufficiently accurate to compute the main formulas (6), (10), (12), (13) up to $\sim$.

The Abel-Jacobi map is an isomorphism (as additive groups) from $E_\epsilon(\mathbb{C})$ to $\mathbb{C}/\Lambda$. Identify $E_\epsilon(\mathbb{C})/\pm$ with $\mathbb{P}^1(\mathbb{C})$ using $\pm P \mapsto x(P)$. Let $W := (\mathbb{C}/\Lambda)/\pm$. The Abel-Jacobi map (up to $\pm$) is a bijection:

$$\Psi : \mathbb{P}^1(\mathbb{C}) \to W, \quad \text{where} \quad \Psi(x_0) = \pm\left(\int_{x_0}^{\infty} \frac{\mathrm{d}x}{y} + \Lambda\right). \tag{7}$$

Its inverse is the Weierstrass $\wp$ function [11, 1,II,5,§1].

Each element of W can be written uniquely as

$$\pm(r_1\omega_1 + r_2\omega_2 + \Lambda), \text{ with } r_1 \in [0, 1),\ r_2 \in \left[0, \frac{1}{2}\right], \text{ and if } r_2 \in \left\{0, \frac{1}{2}\right\} \text{ then } r_1 \in \left[0, \frac{1}{2}\right]. \tag{8}$$

Although W is not a group, it inherits the multiplication by N map from $\mathbb{C}/\Lambda$. The order of the element (8) is N if and only if $r_1, r_2 \in \mathbb{Q}$ and the least common multiple of their denominators is N. The image of $\mathbb{P}^1(\mathbb{R})$ under Ψ is a rectangle in W whose corners are the points of order 1 and 2.

Like in the modular description in Section A.1, define the *Cartan* as $C(N) := \{0, \ldots, \lfloor N/2 \rfloor\}$. Let $W(N) \subset W$ be the set of elements of order N, and for each $c \in C(N)$ let $W_c(N) \subseteq W(N)$ be the subset where $r_2 = c/N \in [0, 1/2]$. Let $\mathbf{n}_c(N) := |W_c(N)|$ denote the cardinality of $W_c(N)$.

- Case $c = 0$. Then $\mathbf{n}_0(2) = 1$ and $\mathbf{n}_0(N) = \varphi(N)/2$ for $N > 2$.
- Case $0 < c < N/2$. Then $\mathbf{n}_c(N) = \varphi(d)N/d$, where $d = \gcd(c, N)$.

- Case $c = N/2$. Then $\mathbf{n}_1(2) = 2$ and $\mathbf{n}_{N/2}(N) = \varphi\left(\frac{N}{2}\right)$ for even $N > 2$.

For later use we define $\mathbf{e}_c(N), \mathbf{f}_c(N)$ with these formulas: $\mathbf{n}_c(N) = \mathbf{e}_c(N) \cdot \mathbf{f}_c(N)$ where $\mathbf{e}_2(4) := 1$ and $\mathbf{e}_c(N) := N/d$ otherwise. Define $C_c(N) := \Psi^{-1}(W_c(N)) \subset \mathbb{P}^1(\mathbb{C})$ so that

$$\bigcup_{c \in C} C_c(N) = \Psi^{-1}(W(N)) = \left\{ x(P) \mid P \in E_\epsilon(\mathbb{C}) \text{ has exact order } N \right\}.$$

To prove the main theorem in Sect. 4 it suffices to compute these $x(P)$'s up to $\sim$. We find $C_0(2) = \{1\}$ and $C_1(2) = \{0, \epsilon\}$ from the definition. Next we compute $C_0(N)$ up to $\sim$ for $N \geq 2$. For $C_0(N)$ we have $r_2 = 0$ and $x(P) \in [1, \infty)$. Let $y_1 = x\sqrt{x-1}$. If $\epsilon \ll |x|$ then $y \approx y_1$. For any $x_0 \in [1, \infty)$ we have

$$\Psi(x_0) = \int_{x_0}^{\infty} \frac{\mathrm{d}x}{y} \approx \int_{x_0}^{\infty} \frac{\mathrm{d}x}{y_1} = \pi - 2\arctan\left(\sqrt{x_0 - 1}\right). \tag{9}$$

Equating (9) to $r_1\omega_1 + 0\omega_2$ gives $\pi - 2\arctan\left(\sqrt{x_0-1}\right) \approx 2\pi r_1$. Solving that for x_0 and applying some trigonometry gives $x_0 \sim \sin(\pi r_1)^{-2} = 2/(1 - \cos(2\pi r_1))$. Substituting $r_1 = a/N$ gives

$$C_0(N) \sim \left\{ \sin\left(\frac{a\pi}{N}\right)^{-2} \,\middle|\, 0 < a \leq \frac{N}{2},\ \gcd(a, N) = 1 \right\}. \tag{10}$$

For $C_{N/2}(N)$ we have $r_2 = 1/2$ and $x(P) \in [0, \epsilon]$. Let $y_0 = \sqrt{x(x-\epsilon)(-1)}$. If $|x| \ll 1$ then $y \approx y_0$. Let $x_0 \in [0, \epsilon]$. Working mod $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$, see (7) and (4),(5), we have

$$\Psi(x_0) = \int_{-\infty}^{x_0} \frac{\mathrm{d}x}{y} = \frac{\omega_2}{2} + \int_{0}^{x_0} \frac{\mathrm{d}x}{y} \approx \frac{\omega_2}{2} + \int_{0}^{x_0} \frac{\mathrm{d}x}{y_0} = \frac{\pi + \omega_2}{2} - \arcsin\left(1 - \frac{2x_0}{\epsilon}\right). \tag{11}$$

Equating (11) to $r_1\omega_1 + \frac{1}{2}\omega_2$ gives $x_0 \sim \epsilon \cdot \sin(\pi r_1)^2$. Substituting $r_1 = a/N$ gives

$$C_{\frac{N}{2}}(N) \sim \left\{ \epsilon \cdot \sin\left(\frac{a\pi}{N}\right)^2 \,\middle|\, 0 \leq a \leq \frac{N}{2},\ \gcd\left(a, \frac{N}{2}\right) = 1 \right\}. \tag{12}$$

Now let $r_2 \in (0, 1/2)$ which corresponds to $\epsilon \ll |x_0| \ll 1$ under $\epsilon \to 0^+$. By equating the right-hand side of (9), or that of (11), to $r_1\omega_1 + r_2\omega_2$ and computing a series expansion we find

$$x_0 \sim -4e^{-2\pi i r_1}\left(\frac{\epsilon}{16}\right)^{2r_2}.$$

Substituting $r_1 = -a/N$ (the minus sign does not affect (13)) and $r_2 = c/N$ gives

$$C_c(N) \sim \left\{ -4\,\zeta_N^a \left(\frac{\epsilon}{16}\right)^{\frac{2c}{N}} \,\middle|\, 0 \leq a < N,\ \gcd(a, c, N) = 1 \right\}. \tag{13}$$

Table 1 Summary of notation from Sects. 2.1 and 2.2

Notation	Brief definition	References
p	A Puiseux expansion above $s = 0$	p. 2
v_p	Discrete valuation associated to p	p. 2
$\boldsymbol{e_p}$	Smallest e with $p \in \mathbb{C}\left(\left(s^{1/e}\right)\right)$ $(= v_p(s))$	p. 2
k_p	Residue field associated to p	p. 2
$\boldsymbol{f_p}$	$\left[k_p : \mathbb{Q}\right]$	p. 2
$\boldsymbol{n_p}$	$\boldsymbol{e_p \cdot f_p = \left[\mathbb{Q}((s))[p] : \mathbb{Q}((s))\right]}$	p. 2
$l_s(p)$	Dominant term of p	p. 2
E_ϵ	$y^2 = x(x-\epsilon)(x-1)$, with $0 < \epsilon \ll 1$	p. 3
ω_1, ω_2	Periods of E_ϵ	p. 3
W	$(\mathbb{C}/\Lambda)/\pm$ where $\Lambda = \mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$	p. 3
$W(N)$	Elements of order N in W	p. 3
$C(N)$	The Cartan, $\{0, 1, \ldots, \lfloor N/2 \rfloor\}$	p. 3
$W_c(N)$	c^{th} subset† of $W(N)$ $(c \in C(N))$	p. 3
n_c	$\lvert W_c(N)\rvert = \lvert C_c(N)\rvert = e_c \cdot f_c$	p. 3
$C_c(N)$	c^{th} subset of $\subset \{x(P) \mid P \text{ order } N\}$	p. 4

The a and c in (13) are the a and c appearing in the vectors in Appendix A.1. After rewriting ϵ in terms of s from Sect. 4, Eqs. (10), (12), (13) determine the Galois action.

2.3 Division polynomials

Let K be a field of characteristic 0 and take $a, b \in K$ for which

$$E: \; y^2 = x^3 + ax + b \tag{14}$$

defines an elliptic curve over K. Following [16, Exercise 3.7], the division polynomials $Q_k \in \mathbb{Z}[x, y, a, b]$, $k = 1, 2, \ldots$ are defined by

$$Q_1 := 1, \quad Q_2 := 2y = 2\sqrt{x^3 + ax + b}, \quad Q_3 := 3x^4 + 6ax^2 + 12bx - a^2,$$
$$Q_4 := 4y(x^6 + 5ax^4 + 20bx^3 - 5a^2x^2 - 4abx - 8b^2 - a^3),$$

and the recursion relations [1]

$$Q_{2k+1} = Q_{k+2}Q_k^3 - Q_{k-1}Q_{k+1}^3 \quad \text{for } k \geq 2 \tag{15}$$
$$Q_{2k} = \frac{(Q_{k+2}Q_{k-1}^2 - Q_{k-2}Q_{k+1}^2)Q_k}{Q_2} \quad \text{for } k \geq 3. \tag{16}$$

Recursively define q_k to be Q_k divided all q_d with $d|k$ and $d < k$, so that $Q_k = \prod_{d|k} q_d$. One has $q_1 = 1$, $q_2 = Q_2$, $q_3 = Q_3$, $q_4 = Q_4/Q_2$, and so forth.

Division polynomials have the following properties:

(1) Q_k is in $\mathbb{Z}[x, a, b]$ when k is odd, and in $q_2 \cdot \mathbb{Z}[x, a, b]$ when k is even.
(2) Let $\mathcal{O}$ be the identity in $(E(\overline{K}), +)$, let $E[k]$ be the points P in $E(\overline{K})$ with $kP = \mathcal{O}$. Then Q_k has one pole, of order $k^2 - 1$ at $\mathcal{O}$, and a root of order 1 at every $P \in E[k] - \{\mathcal{O}\}$. The roots of q_k are the points of exact order k, denoted $E[=k] \subseteq E[k]$.
(3) The $\pm$ below means: choose only one element of each pair $\{P, -P\} \subset E[k]$ (this is not relevant for $k = 2$ because $P = -P$ when $P \in E[2]$).

[1]† These subsets are Galois orbits, see Sect. 4, or alternatively, Appendix A.1

$$\text{If } k \text{ is odd}: \quad Q_k = k \prod_{P \in (E[k]-\{\mathcal{O}\})/\pm} (x - x(P)) \in \mathbb{Q}[x, a, b]. \tag{17}$$

$$\text{If } k \text{ is even}: \quad Q_k/y = k \prod_{P \in (E[k]-E[2])/\pm} (x - x(P)) \in \mathbb{Q}[x, a, b].$$

$$\text{If } k > 2: \quad q_k = a_k \prod_{P \in E[=k]/\pm} (x - x(P)) \in \mathbb{Q}[x, a, b], \tag{18}$$

where $a_k = p$ if k is a power of the prime p and 1 otherwise.

$$\text{For } k = 2: \quad q_2^2 = 4 \prod_{P \in E[=2]} (x - x(P)) \in \mathbb{Q}[x, a, b].$$

The formulas imply that Q_k is square-free, and if $d|k$, then $Q_d|Q_k$. Let m_k denote the number of elements of $E[=k]$. We have $m_2 = 3$, $m_3 = 8$, and $12|m_k$ when $k > 3$. Note that $\deg_x(q_k) = m_k/2$.

3 Equations for $X_1(N)$ and modular units

The definitions of Q_k and q_k are not completely canonical; recurrence relations (15) and (16) are preserved under *scaling*. Scaling means multiplying Q_k by α^{k^2-1} and q_k by α^{m_k} for some fixed $\alpha \neq 0$. To obtain expressions that are independent of scaling, we take quotients

$$\tilde{Q}_k = \frac{Q_k}{q_2^{(k^2-1)/3}} \quad \text{and} \quad \tilde{q}_k = \frac{q_k}{q_2^{m_k/3}}. \tag{19}$$

As before we have $\tilde{Q}_k = \prod_{d|k} \tilde{q}_d$. Since $\tilde{Q}_k = \tilde{q}_k = 1$ for $k \in \{1, 2\}$, we have $\tilde{Q}_k = \tilde{q}_k$ for $k < 6$. To avoid the fractional exponent in (19), we also introduce

$$F_3 = \tilde{q}_3^3 = \frac{q_3^3}{q_2^8} \quad \text{and} \quad F_k = \tilde{q}_k \text{ for } k > 3. \tag{20}$$

Let $\tilde{Q}_{k\backslash 3}$ be $\tilde{Q}_k/\tilde{q}_3$ if $3|k$ and $\tilde{Q}_k$ otherwise. Because $\tilde{Q}_k$ comes from Q_k by scaling, it satisfies the recurrence relations. These relations inductively show that

$$\tilde{Q}_{k\backslash 3} = \prod_{3 \neq d|k} \tilde{q}_d \in \mathbb{Z}[F_3, F_4]. \tag{21}$$

Assuming that $F_3, F_4 \in \mathbb{Q}(x, a, b)$ are algebraically independent over $\mathbb{Q}$, Appendix B shows (21), and that $\tilde{Q}_{k\backslash 3}$ is *primitive* in $\mathbb{Z}[F_3, F_4]$, i.e. the gcd of the coefficients in $\mathbb{Z}$ is 1. The product in (21) is square-free since (17) and (18) are square-free. Then by induction $F_4, F_5, F_6, \ldots$ from (20) are primitive, co-prime, and square-free in $\mathbb{Z}[F_3, F_4]$.

Example 2 To illustrate the various expressions defined above, take E from Eq. (14). We substitute $a = 0$ and $b = 1$ to reduce expression sizes, so that $E : y^2 = x^3 + 1$. The $\mathbb{Q}$-rational torsion subgroup of this curve is isomorphic to $\mathbb{Z}/6\mathbb{Z}$. Specifically, it has the 2-torsion point $(-1, 0)$, the 3-torsion points $(0, \pm 1)$, and the 6-torsion points $(2, \pm 3)$. For $k > 2$, a point P on E has order *equal* to k if and only if $x(P)$ is a root of q_k, or equivalently, a root of F_k. A similar statement holds for Q_k if one replaces "equal to k" with "divides k".

Table 2 Notation related to division polynomials

E	$y^2 = x^3 + ax + b, a = -3j_0, b = -2j_0$	pp. 5, 7
j_0, j	$j_0 = j/(j-1728)$, j = j-invariant of E	p. 6
s	$s = 1/j$, roots(s) = {cusps of $X_1(N)$}	p. 8
$E[=k]$	{points on E of exact order k}	p. 5
m_k	m_k = # points of exact order k	p. 6
Q_k	division polynomial of E	pp. 5, 6
q_k	roots(q_k) = {$x(P)$ | P has order k},$Q_k = \prod_{d\|k} q_k$	pp. 5, 6
$\tilde{Q}_k$	rescaling of Q_k to make it unique, $\tilde{Q}_k = Q_k/Q_2^{(k^2-1)/3}$	p. 6
$\tilde{q}_k$	rescaling of q_k, $\tilde{Q}_k = \prod_{d\|k} \tilde{q}_d$	p. 6
$\{F_k\}$	basis of modular units, $F_k \in \mathbb{Q}(x, j_0) = \mathbb{Q}(x, j) = \mathbb{Q}(x, s)$	[6, 18]
F_2, F_3	$F_2 = q_2^4/\left(1728j_0^2(j_0-1)\right)$, $F_3 = \tilde{q}_3^3 = q_3^3/q_2^8$	pp. 6, 7
$F_k, k > 3$	$F_k = \tilde{q}_k = q_k/q_2^{m_k/3}$	page 6

A point P has order 2 if and only $x(P)$ is a root of q_2^2, or equivalently, a root of F_2 given in the next section. We compute

$$Q_2 = 2y,\ \ Q_2^2 = q_2^2 = 4(x^3+1),\ \ Q_3 = q_3 = 3x\left(x^3+4\right),\ \ Q_4 = q_2q_4 = 4y\left(x^6+20x^3-8\right),$$
$$\tilde{Q}_1 = \tilde{q}_1 = \tilde{Q}_2 = \tilde{q}_2 = \tilde{Q}_{3\backslash 3} = 1,\ \ \tilde{Q}_3 = \tilde{q}_3 = \frac{3x\left(x^3+4\right)}{\left(4(x^3+1)\right)^{\frac{4}{3}}}.$$
$$F_3 = \tilde{q}_3^3 = \frac{\left(3x\left(x^3+4\right)\right)^3}{\left(4(x^3+1)\right)^4},\ \ \tilde{Q}_4 = \tilde{q}_4 = F_4 = \frac{2\left(x^6+20x^3-8\right)}{\left(4(x^3+1)\right)^2}.$$

For a more interesting example of $\tilde{Q}_{k\backslash 3}$, we find

$$q_6 = \left(x^3-8\right)\left(x^9+228x^6+48x^3+64\right),\ \text{so}\ \tilde{Q}_{6\backslash 3}$$
$$= \tilde{q}_1\tilde{q}_2\tilde{q}_6 = \tilde{q}_6 = q_6/q_2^8 = F_6 = -(F_4^2 + F_3 - F_4).$$

Our main curve below has coefficients in $\mathbb{Q}(j)$ instead of $\mathbb{Q}$, so its Q_k, q_k, etc., will be larger.

Henceforth, E will be the curve

$$E : y^2 = x^3 - 3j_0x - 2j_0, \quad \text{where } j_0 := \frac{j}{j-1728}. \tag{22}$$

Now E is defined over $K = \mathbb{Q}(j)$, where j is transcendental over $\mathbb{Q}$. So a and b from Sect. 2.3 will be $-3j_0$ and $-2j_0$ from here on. Now $q_3, q_4, \ldots$ are in $\mathbb{Q}[x, j_0]$ (also in $\mathbb{Z}[x, j_0]$ but we do not use that). The j-invariant of E is j, and the j-invariant of E_ϵ is

$$\frac{2^8(\epsilon^2-\epsilon+1)^3}{\epsilon^2(\epsilon-1)^2}. \tag{23}$$

In Sect. 4 we will equate (23) to j in order to translate formulas computed in terms of ϵ in Sect. 2.2 to similar formulas for E.

Up to a simple transformation, E is the universal elliptic curve E_j from Diamond and Shurman's book [9]. Sections 7.5 and 7.7 in [9] show that the modular curve $X_1(N)$ can be represented with the equation q_N when $N > 2$. In particular, q_N is irreducible in $\mathbb{Q}[x, j_0]$. Likewise F_N is irreducible in $\mathbb{Z}[F_3, F_4]$. Although $q_2 \notin \mathbb{Q}[x, j_0]$, its square $4(x^3 - 3j_0x - 2j_0)$ is an equation for $X_1(2)$ that lies in $\mathbb{Q}[x, j_0]$.

Table 3 $l_s(p+1)$ **for one** p **from each conjugacy class** C_c **over** $\mathbb{Q}((s))$

	$l_s(p+1)$ with $p \in C_0$	$l_s(p+1)$ $p \in C_1$	$l_s(p+1)$ $p \in C_2$	$l_s(p+1)$ $p \in C_3$	$l_s(p+1)$ $p \in C_4$
q_2^2	3	$-24s^{1/2}$			
q_3	4	$-12s^{1/3}$			
q_4	6	$-12s^{1/4}$	$0\,s^{1/2} - 672s$		
q_5	$3\sin(\pi/5)^{-2}$	$-12s^{1/5}$	$-12s^{2/5}$		
q_6	12	$-12s^{1/6}$	$-12(-s)^{1/3}$	$-12s^{1/2}$	
q_7	$3\sin(\pi/7)^{-2}$	$-12s^{1/7}$	$-12s^{2/7}$	$-12s^{3/7}$	
q_8	$3\sin(\pi/8)^{-2}$	$-12s^{1/8}$	$-12(-s)^{1/4}$	$-12s^{3/8}$	$-12(2s)^{1/2}$
q_9	$3\sin(\pi/9)^{-2}$	$-12s^{1/9}$	$-12s^{2/9}$	$-12(\zeta_3 \cdot s)^{1/3}$	$-12s^{4/9}$

4 The valuation of a division polynomial at a cusp

Recall from Sect. 2.1 that a *place* on $X_1(N)/\mathbb{Q}$ is a discrete valuation $v_P : \mathbb{Q}(X_1(N)) \to \mathbb{Z} \bigcup\{\infty\}$. Such a place is a *cusp over* $\mathbb{Q}$ when $v_P(j) < 0$. A function $g \in \mathbb{Q}(X_1(N))$ is called a *modular unit* if every place with $v_P(g) \neq 0$ is a cusp. If $1 < k \neq N > 2$ then F_k is a *modular unit* on $X_1(N)$, see [6, Section 2]. For $k > 3$ this F_k equals $\tilde{q}_k$ which is a rescaled version of q_k (a defining equation for $X_1(k)$ and the main factor of the division polynomial Q_k). For $k = 3$ we need to take the cube of $\tilde{q}_3$ to obtain a modular unit F_3. To construct a modular unit from q_2, we need to take its 4^{th} power and scale it to

$$F_2 = \frac{q_2^4}{1728 j_0^2 (j_0 - 1)}. \tag{24}$$

Let $s = 1/j$. Translating the definition given above, a *cusp over* $\mathbb{Q}$ is a place above $s = 0$, which corresponds to a *conjugacy class of Puiseux expansions at* $s = 0$, see Sect. 2.1. Conjugation is always over $\mathbb{Q}((s))$ in this paper.

From (15) and (16) one can compute $Q_2, Q_3, \ldots$ and then $q_2^2, q_3, q_4, \ldots \in \mathbb{Q}[x, j_0] \subset \mathbb{Q}(s)[x]$. We computed Puiseux expansions of q_N (or q_2^2 if $N = 2$) at $s = 0$ for $N \leq 9$. Newton's algorithm gives arbitrarily many terms, but only dominant terms will be needed. Table 3 lists the dominant term of $\mathbf{p}+1$ for *one* Puiseux expansion $\mathbf{p}$ *from each* conjugacy class $\{\mathbf{p}, \ldots\}$ (which has $\mathbf{n_p} := \mathbf{e_p}\,\mathbf{f_p}$ elements as we will see in the observations below). Notice that in row q_4, column C_2 the term $0s^{1/2}$ predicted by the formula in Observation (4) below vanishes.

We use $\mathbf{p} + 1$ and $x + 1$ rather than $\mathbf{p}$ and x because when $j \to \infty$ the curve E in (22) becomes singular at $x = -1$. This is in contrast to E_ϵ which becomes singular at $x = 0$ when $\epsilon \to 0$.

The equation for $X_1(2)$ is $q_2^2 = 4(x^3 - 3j_0 x - 2j_0)$, where $j_0 = j/(j-1728) = 1/(1-1728s)$. To illustrate Table 3 for $N = 2$, factor $q_2^2 = 4(x - \mathbf{p}_0)(x - \mathbf{p}_{1a})(x - \mathbf{p}_{1b}) \in \overline{\mathbb{Q}((s))}[x]$. Row q_2^2 in Table 3 gives $l_s(\mathbf{p}_0 + 1) = 3$, $l_s(\mathbf{p}_{1a} + 1) = -24s^{1/2}$, and its conjugate $l_s(\mathbf{p}_{1b} + 1) = 24s^{1/2}$. This means

$$q_2^2 = 4((x+1) - 3 + \cdots)\left((x+1) + 24s^{1/2} + \cdots\right)\left((x+1) - 24s^{1/2} + \cdots\right), \tag{25}$$

where the dots indicate terms with higher powers of s. Likewise, for $N > 2$,

$$q_N = a_N \prod_{c=0}^{\lfloor N/2 \rfloor} \left((x+1) - (\mathbf{p}_{c,*} + 1)\right) = a_N \prod_{c=0}^{\lfloor N/2 \rfloor} \left(x - \mathbf{p}_{c,*}\right),$$

where $l_s\left(\mathbf{p}_{c,*} + 1\right)$ are the conjugates of the term listed in row q_N, column C_c.

Example 3 Counting conjugates, row q_8 in Table 3 gives two $\mathbf{p}$'s with $v_s(\mathbf{p}+1)=0$, eight $\mathbf{p}$'s with $v_s(\mathbf{p}+1)=1/8$, four with $1/4$, eight with $3/8$, and two with $1/2$. Indeed, $\deg_x(q_8)=m_8/2$ equals $2+8+4+8+2$.

Now take as an example the conjugacy class C_1 (a *cusp over* $\mathbb{Q}$) on $X_1(3)$. This cusp is a *place* on $X_1(3)/\mathbb{Q}$ and can be represented with a Puiseux expansion $\mathbf{p}$ as in Sect. 2.1. Row q_3, column C_1 gives $\mathbf{p}+1=-12s^{1/3}+\cdots$. Viewing q_8 as an element of $\mathbb{Q}(X_1(3))=\mathbb{Q}(s)[x]/(q_3)$, we can insert this data into Eq. (2) to find

$$v_{\mathbf{p}}(q_8)=3\cdot\left(2\min\left(\frac{1}{3},0\right)+8\min\left(\frac{1}{3},\frac{1}{8}\right)+4\min\left(\frac{1}{3},\frac{1}{4}\right)+8\min\left(\frac{1}{3},\frac{3}{8}\right)+2\min\left(\frac{1}{3},\frac{1}{2}\right)\right) \tag{26}$$

(the $+1$'s cancelled out). Omitting the factor of 3 gives the *unweighted order* from Sect. 2.1.

Table 3 was obtained by computing Puiseux expansions of q_N for $N\le 9$. Let $k,N>1$ and $k\neq N$. Example 3 shows how one can use Table 3 to compute the valuation of q_k (or q_2^2 if $k=2$) at any cusp of $X_1(N)$ when $k,N\le 9$. To obtain a general formula, we will show that Observations (1)–(6) below, which hold in Table 3, hold for all $N>1$.

(1) q_N (q_2^2 if $N=2$) has $\lfloor N/2\rfloor+1$ conjugacy classes (a.k.a. Galois orbits) $C_0,C_1,\ldots,C_{\lfloor N/2\rfloor}$ of Puiseux expansions at $s=0$. We number them so that if $\mathbf{p}\in C_c$ then $v_s(\mathbf{p}+1)=c/N$ except when $(N,c)=(4,2)$. This unique exceptional case is the irregular cusp of $X_1(4)$, where the $s^{c/N}$ term in Table 3 is $0s^{1/2}$ and $v_s(\mathbf{p}+1)=1$ instead.
(2) C_0 has $l_s(\mathbf{p}+1)=12/\left(2-\zeta_N-\zeta_N^{-1}\right)=3\sin(\pi/N)^{-2}$. The residue field is $\mathbb{Q}\left(\zeta_N+\zeta_N^{-1}\right)$.
(3) If $0<c<N/2$, then $\mathbf{p}\in C_c$ has $l_s(\mathbf{p}+1)=-12\,(\zeta_d\cdot s)^{c/N}$ (always up to conjugation) with $d=\gcd(c,N)$ and residue field $\mathbb{Q}(\zeta_d)$.
(4) If $N\neq 4$ is even, then $\mathbf{p}\in C_{N/2}$ has

$$l_s(\mathbf{p}+1)=-24s^{1/2}+3\sin^2(\pi/N)\cdot 16s^{1/2}=-12(\beta\cdot s)^{1/2},$$

where $\beta:=\left(\zeta_N+\zeta_N^{-1}\right)^2$ ($\beta\neq 0$ if $N\neq 4$). The residue field is $\mathbb{Q}(\beta)$ (recall Example 1).
(5) $C_c\subset\mathbb{C}\left(\left(s^{1/\mathbf{e}_c(N)}\right)\right)$ has precisely $\mathbf{n}_c(N)$ elements, and the residue field has degree $\mathbf{f}_c(N)$ with $\mathbf{n}_c$, $\mathbf{e}_c$, $\mathbf{f}_c$ as in Sect. 2.2.
(6) Every $\mathbf{p}\in\bigcup_{N,c}C_c(N)$ has a unique $l_s(\mathbf{p}+1)$, so Eq. (2) holds for all combinations. This implies that Example 3 generalizes to Theorem 1 below.

To see why Observations (1)–(6) hold, note that the curve E_ϵ in Sect. 2.2 differs from E by the transformation

$$T:x\mapsto \mathbf{p}_{1a}+(\mathbf{p}_0-\mathbf{p}_{1a})x=(-1-24s^{1/2}+\cdots)+(3+24s^{1/2}+\cdots)x$$

that sends $0,\epsilon$, and 1 to $\mathbf{p}_{1a}$, $\mathbf{p}_{1b}$, and $\mathbf{p}_0$, respectively. From $T(\epsilon)=\mathbf{p}_{1b}$ we find $\epsilon=16s^{1/2}+\cdots$ which can also be computed by equating $j=1/s$ to (23). Section 2.2 gives the ϵ-dominant terms. Substituting $\epsilon\sim 16s^{1/2}$ and applying T yields $l_s(\mathbf{p}+1)$ for every Puiseux expansion of q_N. Observation (6) immediately follows from this, but then Lemma 1 shows that $\mathbf{e}_{\mathbf{p}}$ and $k_{\mathbf{p}}$ can be read from $l_s(\mathbf{p}+1)$, and the remaining observations follow.

Before stating the MinFormula, we need a definition and a recollection. For $t \in [0, 1/2]$, we define the following *unweighted order functions*. For $k = 2$, define $v_2(t) = 4t - 1$, and for $k > 2$, define

$$v_k(t) = s_k \cdot \left(-\frac{m_k}{3} t + \sum_{j=1}^{\lfloor k/2 \rfloor} \mathbf{n}_j(k) \min\left(t, \frac{j}{k}\right) \right), \tag{27}$$

where $s_3 = 3$ and $s_k = 1$ for $k > 3$. Recall that $C_c(N)$ is a conjugacy class of Puiseux expansions, giving one cusp of $X_1(N)/\mathbb{Q}$, or a Galois orbit with $\mathbf{f}_c(N)$ cusps of $X_1(N)/\overline{\mathbb{Q}}$.

Theorem 1 (MinFormula) *Let* $2 < N \neq k > 1$ *and* $0 \leq c \leq N/2$*, then the order of* F_k*, viewed as element of* $\mathbb{Q}(s)[x]/(q_N) = \mathbb{Q}(X_1(N))$*, at* $C_c(N)$ *is*

$$\text{ord}_{C_c(N)}(F_k) = \boldsymbol{e}_c(N) \cdot v_k\left(\frac{c}{N}\right).$$

If $N = 2$ we cannot directly apply this formula to F_k due to its denominator q_2, but the formula still holds for products where q_2 cancels out, such as $F_2^2 F_3$ and $F_2^{m_k/12} F_k$ for $k > 3$.

Proof Observations (1)–(6) imply that the computation in Example 3 works in general, so

$$\mathbf{e}_c(N) \sum_{j=0}^{\lfloor k/2 \rfloor} \mathbf{n}_j(k) \min\left(\frac{c}{N}, \frac{j}{k}\right) \tag{28}$$

is the order of q_k (or q_2^2 if $k = 2$) at $C_c(N)$ for any $N, k > 1$ with $N \neq k$. Theorem 1 follows by applying Eq. (28) to F_k in Eqs. (20) and (24), simplifying $\min(t, 0) = 0$ and $\min(t, 1/2) = t$, and noting that the denominator $1728 j_0^2 (j_0 - 1)$ in Equation (24) has a root of order 1 at $s = 0$.

Remark 1 A cusp over $\mathbb{Q}$ corresponds to $\mathbf{f}_c(N)$ cusps over $\overline{\mathbb{Q}}$. Since the degree of the divisor of a function is zero,

$$\sum_{c=0}^{\lfloor N/2 \rfloor} \mathbf{f}_c(N)\, \mathbf{e}_c(N)\, v_k\left(\frac{c}{N}\right) = 0.$$

If N is prime, then $\mathbf{e}_c(N)\, \mathbf{f}_c(N) = \mathbf{n}_c(N)$ is N for $c > 0$, and $v_k(0) = 0$, so

$$\sum_{c=0}^{\lfloor N/2 \rfloor} N\, v_k\left(\frac{c}{N}\right) = 0.$$

Letting $N \to \infty$, we see $\int_0^{1/2} v_k(t)\mathrm{d}t = 0$. Since $\int_0^{1/2} (\min(t, (j/k)) - 4(j/k)(1 - (j/k))t)\mathrm{d}t$ equals 0 for any $(j/k) \in [0, 1/2]$, we do not need a formula for m_k, and can instead rewrite Eq. (27) as

$$v_k(t) = s_k \sum_{0<j<k/2} \mathbf{n}_j(k) \left(\min\left(t, \frac{j}{k}\right) - 4\frac{j}{k}\left(1 - \frac{j}{k}\right) t \right), \quad \text{for } k \geq 3. \tag{29}$$

This is the formula implemented in [25]. The sum (29) does not change if one replaces $0 < j < k/2$ by $0 \leq j \leq k/2$ because the summand vanishes at $j \in \{0, k/2\}$. Equation (29) with the factor s_k removed gives the unweighted order function for $\tilde{q}_k$ (recall that $F_3 = \tilde{q}_3^3$ and $F_k = \tilde{q}_k$ if $k > 3$).

5 The degree of F_7/F_8 in $X_1(N)$

In this section we use Theorem 1 to prove an upper bound for the $\mathbb{Q}$-gonality of $X_1(N)$.

Let $\nu(t) := \nu_7(t) - \nu_8(t)$, and let $m(t) := \max(0, \nu(t))$ as in Figure 1. Define

$$B_0(N) := \sum_{0<c<N/2} \mathbf{n}_c(N)\, m\left(\frac{c}{N}\right) \quad \text{and} \quad B_1(N) := \sum_{0<c<N/2} N\, m\left(\frac{c}{N}\right).$$

Theorem 1 gives

$$\mathrm{div}\left(\frac{F_7}{F_8}\right) = \sum_{0<c<N/2} \mathbf{e}_c(N)\, \nu\left(\frac{c}{N}\right) C_c, \quad \mathrm{sodeg}\left(\frac{F_7}{F_8}\right) = B_0(N) \le B_1(N). \tag{30}$$

We omit the terms $c = 0$ and $c = N/2$ in these sums because ν vanishes there. By Eq. (27)

$$\nu(t) = 7\min\left(t, \frac{1}{7}\right) + 7\min\left(t, \frac{2}{7}\right) + 7\min\left(t, \frac{3}{7}\right) - 8\min\left(t, \frac{1}{8}\right) - 4\min\left(t, \frac{1}{4}\right) - 8\min\left(t, \frac{3}{8}\right) - 2t.$$

Lemma 2 *If N is relatively prime to $420 = 3\cdot 4\cdot 5\cdot 7$, then $B_1(N) = [11N^2/840]$. In general, $B_1(N) \le [11N^2/840] + 2$ (where equality implies $7 \mid N$), and $B_0(N) \le [11N^2/840]$.*

Proof Consider the intervals $I_1 := (1/4, 2/7]$, $I_2 := (2/7, 1/3)$, $I_3 := (2/5, 3/7]$, and $I_4 := (3/7, 1/2)$. These intervals partition the support of $m(t)$; see Figure 1. We define the functions $m_1(t) := 4t - 1$, $m_2(t) := 1 - 3t$, $m_3(t) := 5t - 2$, and $m_4(t) := 1 - 2t$. The graphs of the $m_j(t)$ over I_j are exactly the line segments in Figure 1. We see $m(t) = m_j(t)$ if $t \in I_j$ and 0 otherwise.

Our goal is to bound

$$B_1(N) = \sum_{j=1}^{4} \sum_{\substack{c/N \in I_j \\ c\in\mathbb{Z}}} N\, m_j\left(\frac{c}{N}\right). \tag{31}$$

Since $N\, m_j(c/N) \in \mathbb{Z}$, we have $B_1(N) \in \mathbb{Z}$. Note that $B_1(N)$ is a Riemann sum of

$$N^2 \int_0^{1/2} m(t)\mathrm{d}t = N^2 \sum_{j=1}^{4} \int_{I_j} m_j(t)\mathrm{d}t = \frac{11}{840} N^2.$$

Since $m(t)$ is piece-wise linear, any error in $B_1(N)$, viewed as an approximation to this integral, must come from the corners:

$$\frac{1}{4}, \frac{2}{7}, \frac{1}{3}, \frac{2}{5}, \frac{3}{7}, \text{ and } \frac{1}{2}.$$

This error depends only on N modulo $420 = 4 \cdot 7 \cdot 3 \cdot 5$. To demonstrate this, let c_{1a} and c_{1b}, respectively, be the minimum and maximum integer c with $c/N \in I_1$. Then c_{1a} equals

$$\frac{N+4}{4}, \frac{N+3}{4}, \frac{N+2}{4}, \text{ or } \frac{N+1}{4}$$

depending on whether N is, respectively, 0, 1, 2, or 3 mod 4. Likewise, the expression for c_{1b} in terms of N depends only on N mod 7. Considering the intervals I_2, I_3, and I_4, we have a total of $4 \cdot 7 \cdot 3 \cdot 5 = 420$ cases. Hence, the difference between the integral and its Riemann sum $B_1(N)$ depends only on N mod 420.

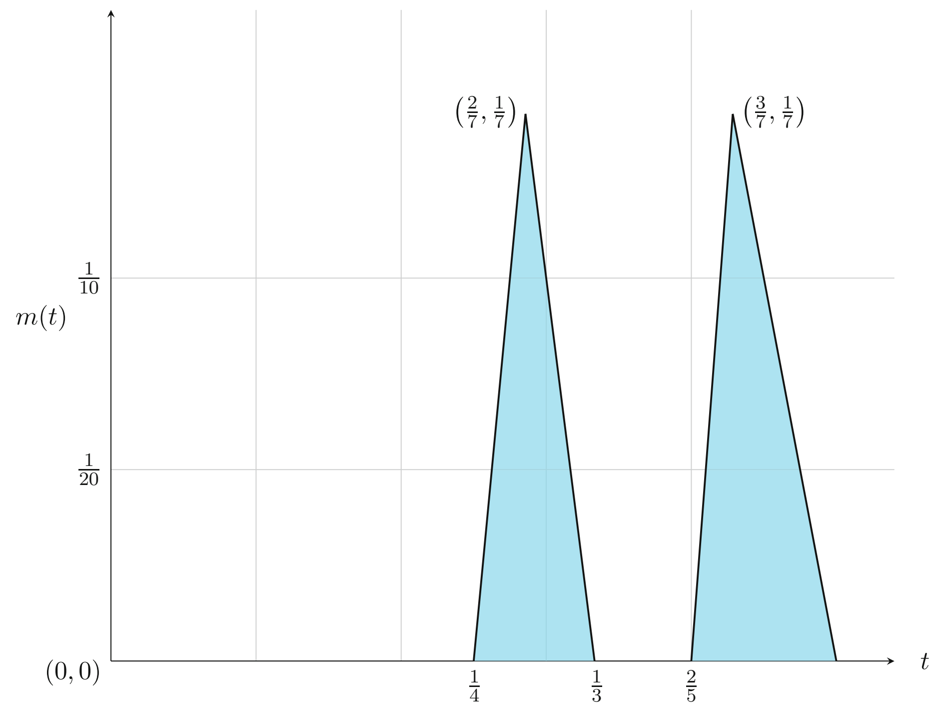

Fig. 1 The function $m(t)$ graphed from 0 to $\frac{1}{2}$

As an example we cover one of the 420 cases, namely $N \equiv 32 \bmod 420$. Here $c_{1a} = (N+4)/4$ and $c_{1b} = (2N-1)/7$, so

$$\sum_{\substack{c/N \in I_1 \\ c \in \mathbb{Z}}} N\, m_1\Big(\frac{c}{N}\Big)$$

has $n := c_{1b} - c_{1a} + 1 = N/28 - 1/7$ terms. The average of these n terms is

$$\frac{1}{2}\Big(N\, m_1\Big(\frac{c_{1a}}{N}\Big) + N\, m_1\Big(\frac{c_{1b}}{N}\Big)\Big) = \frac{15N}{14} + \frac{5}{7},$$

so the $j = 1$ part of $B_1(N)$ in Eq. (31) is $(N/28 - 1/7) \cdot (15N/14 + 5/7)$. Repeating this computation for $j = 2, 3, 4$ and summing, we find

$$B_1(N) = \frac{11N^2}{840} + \frac{43}{105}.$$

Since $|43/105| < 1/2$ we have $B_0(N) \leq B_1(N) = [11N^2/840]$ for any $N \equiv 32 \bmod 420$.

In the same way we calculated the difference between $B_1(N)$ and $\left[11N^2/840\right]$ for all 420 cases, the programs are available at [26]. In all cases with $\gcd(N, 420) = 1$, we found $B_1(N) = \left[11N^2/840\right]$.

If N is prime and $c > 0$, then the factor $\mathbf{n}_c(N)$ in the definition of $B_0(N)$ is N, and hence $B_0(N) = B_1(N)$. So for primes $N > 7$ we find

$$\deg\,(F_7/F_8) = B_0(N) = B_1(N) = \left[11N^2/840\right]. \tag{32}$$

For cases with $\gcd(N, 7) = 1 \neq \gcd(N, 2 \cdot 3 \cdot 5)$ the computation found $B_1(N) \leq \left[11N^2/840\right] + 1$. Moreover, in these cases there is a c/N in some I_j with $\mathbf{n}_c(N) < N$. Then $B_0(N) < B_1(N)$, and so $B_0(N) \leq \left[11N^2/840\right]$.

Table 4 Comparing $\deg\left(F_7/F_8\right) = B_0(N)$ with the $\mathbb{Q}$-gonality for $8 < N \leq 40$

N	$B_0(N)$	Gonality	N	$B_0(N)$	Gonality	N	$B_0(N)$	Gonality	N	$B_0(N)$	Gonality
9	1	1	17	4	4	25	8	5	33	12	10
10	1	1	18	3	2	26	7	6	34	11	10
11	2	2	19	5	5	27	8	6	35	15	12
12	1	1	20	3	3	28	7	6	36	11	8
13	2	2	21	5	4	29	11	11	37	18	18
14	2	2	22	4	4	30	8	6	38	14	12
15	2	2	23	7	7	31	13	12	39	18	14
16	3	2	24	6	4	32	10	8	40	16	12

Table 5 Comparing $B_0(N)$ with the upper bound on the $\mathbb{Q}$-gonality from [6, Table 1]

N	$B_0(N)$	gon≤	N	$B_0(N)$	gon≤	N	$B_0(N)$	gon≤	N	$B_0(N)$	gon≤
41	22	22	51	30	24	61	49	49	71	66	66
42	15	12	52	26	21	62	37	36	72	45	32
43	24	24	53	37	37	63	45	36	73	70	70
44	19	15	54	26	18	64	40	32	74	54	51
45	23	18	55	37	30	65	53	42	75	63	40
46	21	19	56	31	24	66	39	30	76	56	45
47	29	29	57	37	30	67	59	58	77	75	60
48	19	16	58	33	31	68	45	36	78	52	42
49	31	21	59	46	46	69	55	44	79	82	82
50	23	15	60	31	24	70	45	36	80	60	48

For the remaining cases $\gcd(N, 7) \neq 1$ we found $B_1(N) \leq \left[11N^2/840\right] + 2$. The smallest N for which that is sharp is $N = 49$. We check that a multiple of $7/N$ is in each of the intervals $(1/4, 1/3)$ and $(2/5, 1/2)$. These two multiples c/N of $7/N$ each have $\mathbf{n}_c(N) < N$. So $B_0(N) \leq B_1(N) - 2$. Hence we still have $B_0(49) \leq \left[11 \cdot 49^2/840\right]$. The next N with $B_1(N) = \left[11N^2/840\right] + 2$ is $N = 91$. For $N \geq 91$ the intervals $(1/4, 1/3)$ and $(2/5, 1/2)$ each have at least 7 consecutive c/N's, and so $\gcd(c, N) > 1$ (which implies $\mathbf{n}_c(N) < N$) happens at least once in each of those intervals. Then the same argument shows that $B_0(N) \leq \left[11N^2/840\right]$.

Theorem 2 *For $N \neq 7, 8$ the modular unit*

$$\frac{F_7}{F_8} : X_1(N) \to \mathbb{P}^1$$

has degree

$$\deg\left(\frac{F_7}{F_8}\right) = B_0(N) \leq \left[\frac{11N^2}{840}\right],$$

with equality when N is prime. If $N > 8$, then this is an upper bound for the gonality.

Proof The theorem follows from Eq. (30), Lemma 2 and Eq. (32) for prime N. We need $N \neq 7, 8$ to ensure $F_7, F_8 \neq 0$. If F_7/F_8 is not constant (if its degree $B_0(N)$ is not 0) then $B_0(N)$ is a gonality bound. If $N < 7$ then $B_0(N) = 0$. It is easy to check that $B_0(N) > 0$ for $N > 8$.

Tables 4 and 5 show that if N is not prime, then the gonality is usually smaller than $B_0(N)$. If N is prime, then $B_0(N)$ is an excellent gonality bound; the only primes $N < 250$ for which [6, Table 1] gives a sharper bound are 31, 67, 101 where it is only one less than $B_0(N)$.

Acknowledgements

The first author was supported by NSF grants 1618657 and 2007959. The authors would like to extend a very appreciative thanks to the anonymous referees for their careful reading and detailed comments. The second author would also like to thank Katherine E. Stange for the many helpful conversations about modular units.

Author details

[1]Dept. of Mathematics, Florida State University, Tallahassee, FL 32306, USA, [2]Department of Mathematics, University of Connecticut, 341 Mansfield Road U1009, Storrs, CT 06269-1009, USA.

A. A second proof of the MinFormula

A.1 Cusps: a modular interpretation

Take the congruence subgroup

$$\Gamma_1(N) = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in \mathrm{SL}_2(\mathbb{Z}) \;\middle|\; \begin{bmatrix} a & b \\ c & d \end{bmatrix} \equiv \begin{bmatrix} 1 & * \\ 0 & 1 \end{bmatrix} \bmod N \right\}$$

where $*$ indicates the entry is unspecified. The extended complex upper half plane is

$$\overline{\mathcal{H}} = \mathcal{H} \cup \mathbb{Q} \cup \{\infty\},$$

where $\mathcal{H}$ is the usual complex upper half plane. The groups $\Gamma_1(N) \subseteq \mathrm{SL}_2(\mathbb{Z})$ act on the extended complex upper half plane $\overline{\mathcal{H}}$ by fractional linear transformations. The quotient is the modular curve $X_1(N)$.

Following [9, Chapter 3.8] and similar to Sect. 2.2, we represent cusps of $X_1(N)/\overline{\mathbb{Q}}$ with pairs of order N vectors

$$\pm \begin{bmatrix} a \\ c \end{bmatrix} \in (\mathbb{Z}/N\mathbb{Z})^2 .$$

The Galois action on the cusps can be represented with matrices of the form

$$\pm \begin{bmatrix} y & z \\ 0 & 1 \end{bmatrix} \in \mathrm{GL}_2(\mathbb{Z}/N\mathbb{Z})$$

on the order N vectors in $(\mathbb{Z}/N\mathbb{Z})^2$, see [9, Sections 7.6 and 7.7]. Two vectors

$$\begin{bmatrix} a' \\ c' \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a \\ c \end{bmatrix}$$

represent the same cusp when

$$\begin{bmatrix} a' \\ c' \end{bmatrix} = \pm \begin{bmatrix} a + jc \\ c \end{bmatrix}$$

for some $j \in \mathbb{Z}$. Two cusps represented this way are in the same Galois orbit if and only if $c = \pm c'$. Hence Each Galois orbit is uniquely determined by $\pm c$, in other words, by an element of the *Cartan* $C(N) := (\mathbb{Z}/N\mathbb{Z})/\pm$, which is identified with $\{0, \ldots, \lfloor N/2 \rfloor\}$. We will denote such orbit by $C_c(N)$. Let $\mathbf{n}_c(N)$, $\mathbf{e}_c(N)$, $\mathbf{f}_c(N)$ be as in Sect. 2.2. There are $\mathbf{f}_c(N)$ cusps in $C_c(N)$, each of which is represented by $\mathbf{e}_c$ pairs of vectors in $(\mathbb{Z}/N\mathbb{Z})^2$, for a total of $\mathbf{n}_c = \mathbf{e}_c\,\mathbf{f}_c$ pairs.

The *width* of a cusp [9, pp. 59 and 60] is defined as follows. Let $A \in \mathrm{SL}_2(\mathbb{Z})$ be such that $A \cdot \left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right] = \infty = \left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$. The *width* $\mathbf{e}_{\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]}(N)$ is the smallest positive integer for which

$$A \begin{bmatrix} 1 & \mathbf{e}_{\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]}(N) \\ 0 & 1 \end{bmatrix} A^{-1} \in \Gamma_1(N).$$

A computation shows that this is $N/\gcd(c, N)$. Thus the *width* $\mathbf{e}_{\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]}(N)$ is $N/\gcd(c, N)$, which equals the number $\mathbf{e}_c(N)$ from Sects. 2.2 and 4 with *one exception*, namely $C_2(4)$. The cusp corresponding to $\left[\begin{smallmatrix} 1 \\ 2 \end{smallmatrix}\right]$ on $X_1(4)$ is the lone cusp in the orbit $C_2(4)$. It is the only irregular cusp for any modular curve $X_1(N)$, $X_0(N)$, or $X(N)$ [9, p. 75]. It has width 2, but it has 'order' 1. Throughout this paper $\mathbf{e}_c(N)$ denotes the width, except for the case $\mathbf{e}_2(4)$ where it denotes the 'order' which is 1.

5.1 Siegel functions

We would like to define a class of functions on the complex upper half plane $\mathcal{H}$.

Definition 2 Let $(a_1, a_2) \in \mathbb{Q}^2 - \mathbb{Z}^2$. For $\tau \in \mathcal{H}$, define the *Siegel function* associated to (a_1, a_2), denoted $g_{(a_1,a_2)}$, by the product

$$g_{(a_1,a_2)}(\tau) := -q^{\frac{1}{2}\mathbb{B}_2(a_1)} e^{2\pi i \frac{1}{2}(a_2(a_1-1))}(1 - e^{2\pi i a_2} q^{a_1}) \prod_{n=1}^{\infty} (1 - e^{2\pi i a_2} q^{n+a_1})(1 - e^{-2\pi i a_2} q^{n-a_1}),$$

where $q = e^{2\pi i \tau}$, and $\mathbb{B}_2(x) = x^2 - x + \frac{1}{6}$ is the second Bernoulli polynomial.

One can check that adding an integral vector to (a_1, a_2) does not change the order of $g_{(a_1,a_2)}$, so we can interpret (a_1, a_2) as a non-zero element of $(\mathbb{Q}/\mathbb{Z})^2$.

We are interested in the divisors of Siegel functions. From the q-expansion, we see that

$$\mathrm{ord}_\infty\, g_{(a_1,a_2)} = \mathbf{e}_\infty \cdot \frac{1}{2}\, \mathbb{B}_2(a_1).$$

Recall, ∞ denotes the standard prime at infinity given by the equivalence class of $\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right]$ under the action of $\Gamma_1(N)$ and $\mathbf{e}_\infty = 1$ is its width. Consider another cusp of the modular curve $X_1(N)$ that corresponds to the orbit of $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$. Let $A \in \mathrm{SL}_2(\mathbb{Z})$ be a matrix such that

$$A \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} a \\ c \end{bmatrix}.$$

When $g_{(a_1,a_2)}$ is a function on $X_1(N)$, the *order of* $g_{(a_1,a_2)}$ at the cusp corresponding to $\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]$ is

$$\mathrm{ord}_{\left[\begin{smallmatrix} a \\ c \end{smallmatrix}\right]}\left(g_{(a_1,a_2)}\right) = \mathbf{e}_c \cdot \frac{1}{2}\, \mathbb{B}_2\left(\left\{\, [(a_1, a_2) \cdot A]_1 \,\right\}\right), \tag{33}$$

where $\{\cdot\} = \cdot - \lfloor \cdot \rfloor$ denotes the *fractional part* and $[\cdot]_1$ denotes the first entry of the vector. The paper [22] has a concise description of the above for an arbitrary modular curve, but [12, Chapter 2] has a more thorough exposition for $X(N)$; specifically, see the boxed equation on page 40. The reader should note that in [12], Kubert and Lang are considering the $q^{\frac{1}{N}}$ expansion. In the remainder of this paper, we will consider Siegel functions of the form $g_{(0,a)}$, with a a nonzero element of $\mathbb{Q}/\mathbb{Z}$ of order dividing N. Following [18], we write

$$H_k := g_{\left(0, \frac{k}{N}\right)}, \quad \text{with } k \in \mathbb{Z} - N\mathbb{Z}.$$

Caution In [18], Streng considers the modular curve $X^1(N)$, while we have $X_1(N)$. The isomorphism $\Gamma^1(N)\setminus\mathcal{H} \to \Gamma_1(N)\setminus\mathcal{H}$ is given by $\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$. This isomorphism sends $g_{(a,0)}$ to $g_{(0,-a)}$; however, $g_{(0,-a)} = -g_{(0,a)}$.

The *unweighted order of* H_k at $c \in C(N)$ is

$$\operatorname{uord}_c(H_k) = \frac{1}{2}\mathbb{B}_2\left(\left\{c \cdot \frac{k}{N}\right\}\right). \tag{34}$$

Note that $x \mapsto \mathbb{B}_2(\{x\})$ is a continuous function even though $x \mapsto \{x\}$ is not.

5.2 Generators of the modular units

Describing the modular units on a given modular curve has long been a subject of interest. A significant motivation of Kubert and Lang's text [12] is to describe the units of $X(N)$ over $\mathbb{Q}(\zeta_N)$. They show that, with the exception of some 2-torsion elements when N is even, the units are generated by the Siegel functions described above.

Motivated by [6, Conjecture 1], Streng [18] has used similar methods to describe all modular units on $X^1(N)$ over $\mathbb{Q}$. Before stating the result we introduce some of the relevant objects. We start with Tate normal form.

Lemma 3 ([18, Lemma 2.1]) *If E is an elliptic curve over a field K of characteristic 0 (such as the elliptic curves in Equations (14) and (22)) and P is a point on E of order greater than 3 with $x(P) \in K$, then the pair $(E, \pm P)$ is isomorphic to a unique pair of the form*

$$E_T : Y^2 + (1-C)XY - BY = X^3 - BX^2, \quad P = (0,0), \tag{35}$$

where $B, C \in K$ and the discriminant

$$D = B^3(16B^2 + (1-20C-8C^2)B + C(C-1)^3) \neq 0.$$

Further, each pair $B, C \in K$ with $D \neq 0$ satisfying (35) yields an elliptic curve and with a distinguished point P of order greater than 3.

This form E_T is called *Tate normal form*. Let $K = \mathbb{Q}(j)$ and E be as in Equation (22) and let $K_0 = K(x_0)$ where x_0 is transcendental over K. Let

$$P = \left(x_0, \sqrt{x_0^3 - 3j_0x_0 - 2j_0}\right).$$

Sending P to $(0,0)$ and E to Tate normal form with affine linear transformations results in expressions $B, C \in K_0$ (computation at [26]). Identify x_0 with x so that K_0 becomes $\mathbb{Q}(x, j)$. Then $B, C \in \mathbb{Q}(x, j)$ are $B = -F_3$ and $C = -F_4$. Due to the uniqueness of the Tate normal form, it should also be possible to write x, j in terms of B, C, and a computation [26] confirms that. Thus $\mathbb{Q}(B, C) = \mathbb{Q}(x, j)$.

A computation [26] shows $F_2 = B^4/D$. Conjecture 1 in [6], proved by Streng [18], says that for $N > 2$, the modular units in $\mathbb{Q}(X_1(N))$ modulo $\mathbb{Q}^*$ are freely generated by $F_2, \ldots, F_{\lfloor N/2\rfloor+1}$.

Considering the Tate normal form over $\mathbb{Z}[B, C]$, we can look at the k^{th} division polynomial $\psi_{k,E_T}(x, y) \in \mathbb{Z}[B, C][x, y]$. As in [18, Example 2.2], evaluating ψ_{k,E_T} at (0, 0) gives:

$$P_1 := \psi_{1,E_T}(0,0) = 1, \qquad P_2 := \psi_{2,E_T}(0,0) = -B, \qquad P_3 := \psi_{3,E_T}(0,0) = -B^3,$$
$$P_4 := \psi_{4,E_T}(0,0) = CB^5, \qquad P_5 := \psi_{5,E_T}(0,0) = -(C-B)B^8,$$
$$P_6 := \psi_{6,E_T}(0,0) = -B^{12}(C^2 - B + C), \qquad P_7 := \psi_{7,E_T}(0,0) = B^{16}(C^3 - B^2 + BC).$$

A computation shows $P_k = (q_3/q_2^3)^{k^2-1} Q_k$ for $k < 5$. This must then be true for all k since both sequences P_k and Q_k satisfy the recurrence relations (15) and (16), which are preserved under scaling (defined in Sect. 3). From Eq. (19),

$$P_k = \left(\frac{q_3}{q_2^3}\right)^{k^2-1} Q_k = \left(\frac{q_3}{q_2^{8/3}}\right)^{k^2-1} \tilde{Q}_k = (\tilde{q}_3)^{k^2-1} \prod_{d|k} \tilde{q}_d = F_3^{\lfloor k^2/3 \rfloor} \prod_{3<d|k} F_d. \tag{36}$$

In particular, the multiplicative group $\langle D, -B, P_4, \ldots, P_k \rangle$ equals $\langle F_2, \ldots, F_k \rangle$. Streng defined F_k for $k > 3$ to be P_k but where all factors P_j with $j < k$ have been removed; Eq. (36) makes this precise.

Since $\psi_{k,E_T}(P) = 0$ if and only if P has order dividing k, we see $F_k(P) = 0$ if and only if P has exact order k. As mentioned in Sect. 3, the polynomial F_N is a model for the modular curve $X_1(N)$ for $N > 3$. The Tate normal form (35) is only defined for $N > 3$, so [6] used x, j coordinates to construct F_2 and F_3. Rewritten in terms of B, C they are $F_2 = B^4 D^{-1}$ and $F_3 = -B$. We can now state the main result of [18]

Theorem 3 [18, Theorem 1.1], [6, Conjecture 1] *The modular units of $X^1(N)$ are given by $\mathbb{Q}^*$ times the free abelian group on $B, D, F_4, F_5, \ldots, F_{\lfloor N/2 \rfloor+1}$, or equivalently, $F_2, \ldots, F_{\lfloor N/2 \rfloor+1}$.*

Streng gives P_k explicitly in terms of Siegel functions.

Lemma 4 [18, Lemma 3.3] *For all $k \in \mathbb{Z} - N\mathbb{Z}$*

$$P_k = \left(\frac{H_1^2 H_3}{H_2^3}\right)^{k^2-1} \frac{H_k}{H_1} \text{ and } D = \left(\frac{H_1^2 H_3}{H_2^3}\right)^{12} H_1^{12}.$$

Defining $\tilde{H}_k := H_k / H_1^{k^2}$, we get

$$P_k = \left(\frac{\tilde{H}_3}{\tilde{H}_2^3}\right)^{k^2-1} \tilde{H}_k, \;\; F_3 = P_2 = \frac{\tilde{H}_3^3}{\tilde{H}_2^8}, \;\; \text{and} \;\; F_2 = \frac{P_2^4}{D} = \tilde{H}_2^4. \tag{37}$$

Setting $t = c/N$, Eq. (34) gives

$$\text{uord}_c\left(\tilde{H}_k\right) = \text{uord}_c\left(H_k\right) - k^2 \text{uord}_c\left(H_1\right) = \frac{1}{2}\left(\mathbb{B}_2(\{\!\{kt\}\!\}) - k^2 \mathbb{B}_2(\{\!\{t\}\!\})\right). \tag{38}$$

We say that a function $f : [0, 1/2] \to \mathbb{R}$ is *k-piecewise linear* if it is continuous and $f''(t) = 0$ for all $t \notin \frac{1}{k}\mathbb{Z}$. Two k-piecewise linear functions coincide if and only if they have: the same initial value $f(0)$, the same initial slope $f'(0^+)$, and the same change in slope at each $t \in \frac{1}{k}\mathbb{Z}$. These three conditions hold for the right-hand sides of (38) and (39) and thus:

$$\mathrm{uord}_c(\tilde{H}_k) = \frac{1}{2}\left((k^2-k)t - \frac{1}{6}(k^2-1)\right) + k\sum_{0<i<k/2}\left(\min\left(t, \frac{i}{k}\right) - t\right). \tag{39}$$

Applying (39) to F_2 and F_3 in (37) produces the unweighted order functions $v_2(t)$ and $v_3(t)$ in Theorem 1.

To verify $v_k(t)$ for the remaining $k > 3$, let $\tilde{v}_k(t)$ be the unweighted order function of $\tilde{q}_k$, i.e. $\tilde{v}_k(t)$ is the right hand side of Eq. (29) without the factor s_k. So $\tilde{v}_2(t) = 0$, $\tilde{v}_3(t) = \frac{1}{3}v_3(t)$ and $\tilde{v}_k(t) = v_k(t)$ for $k > 3$. The unweighted order function for $\tilde{Q}_k = \prod_{d|k}\tilde{q}_d$ according to Theorem 1 and Remark 1 is

$$\sum_{d|k}\tilde{v}_d(t) = \sum_{d|k}\sum_{0<j<d/2}\mathbf{n}_j(d)m_{j/d}(t) = k\sum_{0<i<k/2}m_{i/k}(t), \tag{40}$$

where $m_a(t) = \min(t, a) - 4a(1-a)t$. We also used that k is the sum of $\mathbf{n}_j(d)$, taken over all $0 < j < d/2$ with $d|k$ and $j/d = i/k$.

Applying (39) to $\tilde{Q}_k = \tilde{H}_k/\tilde{H}_2^{(k^2-1)/3}$ gives the same result. To see this, note that $m_{i/k}(t)$, which contains $\min(t, i/k)$, appears in (40) with the same coefficient k as the coefficient of $\min(t, i/k)$ in (39). The terms $\frac{1}{6}(k^2-1)$ from (39) cancel out for $\tilde{H}_k/\tilde{H}_2^{(k^2-1)/3}$ but then the remaining terms $(\cdots)t$ in (39),(40) must also match by the integral argument from Remark 1. This confirms $\tilde{v}_k(t)$ and thus $v_k(t)$ for the remaining k. This gives a second proof for most (except the case $N|k$, see Lemma 4) of the reformulation of Theorem 1 given in Remark 1.

6 B. Proof of primitivity

Proposition 1 *The kth division polynomial of the Tate normal form, P_k, is primitive in $\mathbb{Z}[B, C]$.*

Proof Order the monomials lexicographically with the following rule

$$B^{n_1}C^{n_2} < B^{m_1}C^{m_2} \text{ when } n_1 < m_1 \text{ or } (n_1 = m_1 \text{ and } n_2 < m_2).$$

If $R \in \mathbb{Q}[B, C]$, let $M(R)$ denote the smallest monomial of R. For example, if $R = 3B^2C^5 + B^3C$, then $M(R) = 3B^2C^5$. A key property is $M(R_1R_2) = M(R_1)M(R_2)$.

Let c_k denote $\lceil k/3 \rceil$. It is enough to prove that

$$M(P_k) = (-1)^{c_k}(-B)^{\lfloor k^2/3 \rfloor}C^{c_k(c_k-1)/2} \tag{41}$$

since it shows that P_k has at least one coefficient equal to ± 1.

We will prove (41) by induction. First, a direct verification shows that (41) holds for $k = 1, 2, 3, 4$. Suppose now that k is even, and write $l = \frac{k}{2}$. Recall the recursion relation (16)

$$P_k = \frac{P_l}{P_2}\left(P_{l+2}P_{l-1}^2 - P_{l-2}P_{l+1}^2\right).$$

the smallest monomial of the first summand $P_{l+2}P_{l-1}^2$ is

$$(-1)^{c_{l+2}+2c_{l-1}}(-B)^{\lfloor (l+2)^2/3 \rfloor + 2\lfloor (l-1)^2/3 \rfloor}C^{c_{l+2}(c_{l+2}-1)/2 + c_{l-1}(c_{l-1}-1)}.$$

For the second summand $-P_{l-2}P_{l+1}^2$ it is

$$(-1)^{c_{l-2}+2c_{l+1}}(-B)^{\lfloor (l-2)^2/3\rfloor+2\lfloor (l+1)^2/3\rfloor}C^{c_{l-2}(c_{l-2}-1)/2+c_{l+1}(c_{l+1}-1)}.$$

When $l \equiv 1 \bmod 3$, the second summand has the smallest monomial, and when $l \equiv 2 \bmod 3$, the first summand has the smallest monomial. When $3 \mid l$, we have to consider the exponent of C. In this case, the first summand is the smallest.

In each case, verifying Eq. (41) is straightforward. For example, when $l \equiv 0 \bmod 3$ we have

$$\lfloor l^2/3\rfloor + \lfloor (l+2)^2/3\rfloor + 2\lfloor (l-1)^2/3\rfloor = \frac{4l^2+3}{3} = \lfloor k^2/3\rfloor + 1,$$

and

$$c_l(c_l-1)/2 + c_{l+2}(c_{l+2}-1)/2 + c_{l-1}(c_{l-1}-1) = \frac{l}{3}\left(\frac{2l-3}{3}\right) = \frac{c_k}{2}(c_k-1).$$

Now suppose k is odd and write $k = 2l+1$. Recall the recursion relation (15)

$$P_k = P_{l+2}P_l^3 - P_{l-1}P_{l+1}^3.$$

For the first summand, the smallest monomial is

$$(-1)^{c_{l+2}+3c_l}(-B)^{\lfloor (l+2)^2/3\rfloor+3\lfloor l^2/3\rfloor}C^{c_{l+2}(c_{l+2}-1)/2+3c_l(c_l-1)/2},$$

and for the second summand it is

$$(-1)^{c_{l-1}+3c_{l+1}}(-B)^{\lfloor (l-1)^2/3\rfloor+3\lfloor (l+1)^2/3\rfloor}C^{c_{l-1}(c_{l-1}-1)/2+3c_{l+1}(c_{l+1}-1)/2}.$$

When $l \equiv 0 \bmod 3$, the second summand has the smaller monomial; when $l \equiv 1 \bmod 3$, considering the exponent of C shows the second summand has the smaller monomial; and when $l \equiv 2 \bmod 3$, the first summand has the smaller monomial.

Verifying Eq. (41) is again straightforward for each case. For example, when $l \equiv 1 \bmod 3$

$$\lfloor (l-1)^2/3\rfloor + 3\lfloor (l+1)^2/3\rfloor = \frac{4l^2+4l+1}{3} = \lfloor k^2/3\rfloor,$$

and

$$c_{l-1}(c_{l-1}-1)/2 + 3c_{l+1}(c_{l+1}-1)/2 = \frac{4l^2-2l-2}{18} = \frac{c_k}{2}(c_k-1).$$

Repeating these computations for the remaining cases proves the proposition.

Recall that $-B = F_3 = \tilde{q}_3^3$ and $-C = F_4$. From (36) we find that $\tilde{Q}_{k\backslash 3}$ from Sect. 3 is $P_k/(-B)^{\lfloor k^2/3\rfloor}$ which is primitive in $\mathbb{Z}[B, C] = \mathbb{Z}[F_3, F_4]$ by Eq. (41).

Received: 14 May 2020 Accepted: 3 February 2021 Published online: 15 March 2021

References

1. Baaziz, H.: Equations for the modular curve $X_1(N)$ and models of elliptic curves with torsion points. Math. Comp. **79**(272), 2371–2386 (2010). https://doi.org/10.1090/S0025-5718-10-02332-X
2. Bourdon, A., Ejder, Ö., Liu, Y., Odumodu, F., Viray, B.: On the level of modular curves that give rise to isolated j-invariants. Adv. Math. **357**, 106824 (2019)
3. Carlucci, P.: Cuspidal divisor class groups of non-split Cartan modular curves. Acta Arith. **187**(4), 301–327 (2019). https://doi.org/10.4064/aa8516-6-2018
4. Chen, Y.-H.: Cuspidal $\mathbb{Q}$-rational torsion subgroup of $J(\Gamma)$ of level P. Taiwanese J. Math. **15**(3), 1305–1323 (2011). https://doi.org/10.11650/twjm/1500406301
5. Csirik, J.A.: The Kernel of the Eistenstein Ideal. J. Numb. Theory **92**, 348–375 (2002)

6. Derickx, M., van Hoeij, M.: Gonality of the modular curve $X_1(N)$. J. Algebra **417**, 52–71 (2014). https://doi.org/10.1016/j.jalgebra.2014.06.026
7. Derickx, M., van Hoeij, M., Zeng, J.: Computing Galois representations and equations for modular curves $X_H(l)$. *arXiv e-prints*, art. arXiv:1312.6819 (2014)
8. Derickx, M., Etropolski, A., van Hoeij, M., Morrow, J.S., Zureick-Brown, D.: Sporadic cubic torsion. *arXiv e-prints*, art. arXiv:2007.13929, July (2020). To appear in Algebra & Number Theory
9. Diamond, F., Shurman, J.: A first course in modular forms. Graduate Texts in Mathematics, vol. 228. Springer-Verlag, New York (2005)
10. DLMF. *NIST Digital Library of Mathematical Functions*. http://dlmf.nist.gov/, Release 1.0.26 of 2020-03-15. URL http://dlmf.nist.gov/. F. W. J. Olver, A. B. Olde Daalhuis, D. W. Lozier, B. I. Schneider, R. F. Boisvert, C. W. Clark, B. R. Miller, B. V. Saunders, H. S. Cohl, and M. A. McClain, eds
11. Hurwitz, A., Courant, R.: Vorlesungen über allgemeine Funktionentheorie und elliptische Funktionen. Interscience Publishers Inc, New York (1944)
12. Kubert, D.S., Lang, S.: Modular Units. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science], vol. 244. Springer-Verlag, New York-Berlin (1981)
13. Mazur, B.: Some comments on elliptic curves over general number fields and Brill-Noether modular varieties. 2014. URL http://people.math.harvard.edu/~mazur/papers/For.Maine.0718.2014(3).pdf
14. Najman, F.: Torsion of rational elliptic curves over cubic fields and sporadic points on $X_1(n)$. Math. Res. Lett. **23**(1), 245–272 (2016). https://doi.org/10.4310/MRL.2016.v23.n1.a12
15. Serre, J.-P.: Local fields, volume 67 of Graduate Texts in Mathematics. Springer-Verlag, New York-Berlin. Translated from the French by Marvin Jay Greenberg (1979)
16. Silverman, J.H.: The arithmetic of elliptic curves, volume 106 of of Graduate Texts in Mathematics. 2nd ed. Springer, Dordrecht (2009) https://doi.org/10.1007/978-0-387-09494-6
17. Smith, H.: Ramification in the Division Fields of Elliptic Curves and an Application to Sporadic Points on Modular Curves. *arXiv e-prints*, art. arXiv:1810.04809 (2018)
18. Streng, M.: Generators of the group of modular units for $\Gamma^1(N)$ over $\mathbb{Q}$. *arXiv e-prints*, art. arXiv:1503.08127, Mar (2015)
19. Sutherland, A.: Torsion subgroups of elliptic curves over number fields. (2012a). URL https://math.mit.edu/~drew/MazursTheoremSubsequentResults.pdf
20. Sutherland, A.V.: Constructing elliptic curves over finite fields with prescribed torsion. Mathematics of Computation, **81**(278):1131–1147 (2012). ISSN 00255718, 10886842. http://www.jstor.org/stable/23267989
21. Sutherland, A., van Hoeij, M.: Defining equations for $X_1(N)$ (2014) https://math.mit.edu/~drew/X1_optcurves.html
22. Sutherland, A.V., Zywina, D.: Modular curves of prime-power level with infinitely many rational points. Algebra Numb. Theory **11**(5), 1199–1229 (2017). https://doi.org/10.2140/ant.2017.11.1199
23. Takagi, T.: Cuspidal class number formula for the modular curves $X_1(p)$. J. Algebra **151**(2), 348–374 (1992). https://doi.org/10.1016/0021-8693(92)90119-7
24. van Hoeij, M.: Low Degree Places on the Modular Curve X1(N). (2013). https://www.math.fsu.edu/~hoeij/files/X1N/LowDegreePlaces
25. van Hoeij, M.: Minformula (in file: cusp divisors program) (2013). www.math.fsu.edu/~hoeij/files/X1N
26. van Hoeij, M., Smith, H.: Computations for gonality bound. April 28 (2020). https://doi.org/10.5281/zenodo.4396300
27. Yang, Y.: Modular units and cuspidal divisor class groups of $X_1(N)$. J. Algebra **322**(2), 514–553 (2009). https://doi.org/10.1016/j.jalgebra.2009.04.012
28. Yang, Y., Yu, J.-D.: Structure of the cuspidal rational torsion subgroup of $J_1(p^n)$. J. Lond. Math. Soc. (2) **82**(1), 203–228 (2010). https://doi.org/10.1112/jlms/jdq013
29. Yoo, H.: The rational cuspidal divisor class group of $X_0(N)$. *arXiv e-prints*, art. arXiv:1908.06411 (2019)
30. Yu, J.: A cuspidal class number formula for the modular curves $X_1(N)$. Math. Ann. **252**(3), 197–216 (1980). https://doi.org/10.1007/BF01420083

Publisher's Note