Optimal independent system for the congruence subgroups $\Gamma_0(p)$ and $\Gamma_0(pq)$

NHAT MINH DOAN, SANG-HYUN KIM, MONG LUNG LANG, AND SER PEOW TAN

Dedicated to Professor Ravindra Kulkarni, on the occasion of his eightieth birthday.

Abstract. Does the congruence subgroup $\Gamma_0(N)$ admit freely independent generators whose Frobenius norms satisfy the growth condition $O(N)$? We answer this question affirmatively in the case when $N = p$ or $N = pq$ for odd primes p and q satisfying $|\sqrt{p} - \sqrt{q}| < \sqrt{2}$. In the special case that $N = p$ or $N = p^2$, we can further require that these generators have 0 or N in their $(2,1)$ components, thus establishing a conjecture of Kulkarni; while doing so, we establish that $\Gamma_0(N)$ admits a special (fundamental) polygon in the sense of Kulkarni, encoded by a generalized Farey sequence with denominators less than $\sqrt{4N/3}$.

1. Introduction

Let us denote the classical modular group as

$$\Gamma := \mathrm{PSL}(2,\mathbb{Z}) = \langle S, R \mid S^2 = R^3 = 1 \rangle,$$

where

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad R = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}.$$

As usual, a matrix A is identified with $-A$ in Γ. We let

$$T := R^{-1}S = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Each subgroup Λ of Γ admits a free product decomposition

$$\Lambda = \mathop{\ast}_i \langle s_i \rangle,$$

where the order of each s_i is either two, three, or infinite. Such a generating set $\{s_i\}$ is called a *freely independent generating set* (or *an independent system*) for Λ. When $[\Gamma : \Lambda] < \infty$, the Reidemeister–Schreier process yields a presentation of Λ, and a suitable simplification of the presentation would yield an independent system for Λ. In the case of the congruence subgroup

$$\Gamma_0(N) := \left\{ A \in \Gamma \colon A \equiv \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \pmod{N} \right\},$$

such a process was given by Rademacher [7] for a prime N, and by Chuman [3] for a general integer N.

Date: June 30, 2023.

2010 *Mathematics Subject Classification.* Primary: 11F06; Secondary: 11B57, 30F35.

Key words and phrases. modular group, congruence subgroup, independent generating system, special polygon, Euler totient function, Farey sequence, hyperbolic geometry.

Kulkarni [6] gave a geometric and much more practical algorithm to find an independent system for every finite index subgroup Λ of Γ. In general, a fundamental domain of Λ equipped with side-pairings yields a generating set for Λ, but this is not guaranteed to be independent. Kulkarni utilized/constructed a certain particularly useful type of fundamental domain (*special polygon*) P^*. He encoded this information as a *Farey symbol*, which is equivalent to a maximal ideal polygon P inside P^* with each side of P suitably labeled. The independent system in the *matrix form* can be read off immediately from the output of his algorithm. See Section 2.

For $\Gamma_0(N)$, Kulkarni's algorithm relies on congruence relations among the denominators of certain rational numbers, which correspond to the cusps of a special polygon. After long and meticulous computations, he noted [6] that for $\Gamma_0(p)$ with p prime:

> "For $p < 100$ we observed that the congruences which need to be satisfied for constructing these Farey symbols can actually be lifted to equalities in natural numbers"

and suggested that this could be true in general. Here, the "congruences" mean that the $(2,1)$ components of the resulting independent generators are 0 modulo p. Our first theorem below establishes his conjecture. Here, $\|g\|_F := \sqrt{\mathrm{tr}(gg^t)}$ denotes the Frobenius norm of a matrix g, and $g_{2,1}$ denotes the $(2,1)$ component of g.

Theorem 1.1. *If $N = p$ or $N = p^2$ for a prime p, then $\Gamma_0(N)$ admits a free product decomposition*

$$\Gamma_0(N) = \mathop{\ast}_{i=1}^{k} \langle g_i \rangle * \langle T \rangle$$

for some matrices g_i such that $(g_i)_{2,1} = N$, such that $|\operatorname{tr} g_i| \leqslant N-2$, and such that $\|g_i\|_F < 2N-1$.

We have an analogous description of an independent system for $\Gamma_0(pq)$ as below with p and q close; this theorem covers all twin primes, for instance.

Theorem 1.2. *If $N = pq$ for odd primes p and q satisfying $|\sqrt{p} - \sqrt{q}| < \sqrt{2}$, then the group $\Gamma_0(N)$ admits a free product decomposition*

$$\Gamma_0(N) = \mathop{\ast}_{i=1}^{k} \langle g_i \rangle * \langle T \rangle$$

for some matrices g_i such that

$$(g_i)_{2,1} = \begin{cases} 2N, & \text{if } i \leqslant |q-p|, \\ N, & \text{if } i > |q-p|, \end{cases}$$

such that $|\operatorname{tr} g_i| \leqslant (g_i)_{2,1} - 2$, and such that $\|g_i\|_F < 2(g_i)_{2,1} - 1$.

In particular, we have linear asymptotic bounds $O(N)$ for the $(2,1)$ components, for the traces and also for the norms of certain freely independent generators of $\Gamma_0(N)$ when $N = p$ or pq as above.

Regarding norms of *possibly non-independent* generators, we can summarize the previously known results as follows. When $N = p$, generators of $\Gamma_0(N)$ with Frobenius norms $O(N)$ is described in [11]. For $N = p^r$, a bound of $O(N^2)$ is given in [5]. For a general integer N, a bound $O_\epsilon(N^{1+\epsilon})$ was recently established [9].

Recall the *Farey sequence* Farey_k *of order* k is the set of rational numbers in $[0,1]$ whose denominators are at most k. Letting $\mathscr{H}$ denote the upper-half plane equipped with the hyperbolic metric, we write

$$\mathrm{Farey}_k^* := \mathrm{Farey}_k \cup \{\infty\} \subseteq \mathbb{Q} \cup \{\infty\} =: \hat{\mathbb{Q}} \subseteq S^1 = \partial\mathscr{H},$$

which is called the *extended Farey sequence of order* k. We say a special polygon P^* for $\Lambda \leqslant \Gamma$ is *optimal* if the maximum value of the denominators of the cusps of P^* is as small as possible, among all choices of special polygons for Λ, and we denote by $m(\Lambda)$ this value; see the beginning of Section 3. Note that Λ may have several different optimal special polygons.

We prove Theorem 1.1 in the process of establishing an efficient bound on the value $m(\Gamma_0(N))$ as below. Proving the optimality of the upper bound $\left\lfloor \sqrt{4N/3} \right\rfloor$ below requires more work and is related to the expressibility of integers as certain quadratic forms in one or two variables. More precisely, let us define an integer N to be *cashew* if N is of the form $sa + tb$ for some nonnegative integers s,t,a,b satisfying

$$s + t > a = \left\lfloor \sqrt{4N/3} \right\rfloor > t \geqslant b \geqslant a - s \geqslant 0.$$

Theorem 1.3. *For an integer $N \geqslant 2$, the following hold.*

(1) We have that $m(\Gamma_0(N)) \geqslant \lfloor \sqrt{N} \rfloor$; moreover, equality holds iff N belongs to the list

$$N \in \{2,3,4,5,7,9,11,13,17,19,25,29,31,37,49,53,67,83,127,173\}.$$

(2) If $N = p$ or $N = p^2$ for a prime p, then $m(\Gamma_0(N)) \leqslant \left\lfloor \sqrt{4N/3} \right\rfloor$.

(3) If $N = pq$ for primes p and q satisfying $|\sqrt{p} - \sqrt{q}| < \sqrt{2}$, then

$$\max(p,q) \leqslant m(\Gamma_0(N)) \leqslant \max\left(\left\lfloor \sqrt{4N/3} \right\rfloor, p, q \right).$$

(4) Assume N is as in part 2, or assume $N > 81 \cdot 83$ is as in part 3. Then we have

$$m(\Gamma_0(N)) = \left\lfloor \sqrt{4N/3} \right\rfloor$$

if and only if N is cashew.

An arithmetic interpretation of the list in part 1 is given in Remark 3.5. In part 4 when $N = pq$, the hypothesis $N > 81 \cdot 83$ implies that $\left\lfloor \sqrt{4N/3} \right\rfloor \geqslant \max(p,q)$.

Let us note a special case of the above theorem.

Corollary 1.4. *Let p be a prime in one of the following forms:*

(1) $p = s^2 + t^2 + st - 2s - 2t$, for some integers s,t with $3 \leqslant s < t < s + 1 + 2\sqrt{s+1}$;

(2) $p = 3s^2 - s - 1$, for some integer $s > 0$.

Then we have that $m(\Gamma_0(p)) = \left\lfloor \sqrt{4p/3} \right\rfloor$.

Proof. For part 1, we let $a := s + t - 2$ and $b := t - 2$. It is easy to see that

$$a \leqslant \sqrt{4p/3} < a + 1$$

from the given conditions. In particular, p is cashew and the conclusion follows. By further requiring that $t := s + 1$, we obtain part 2. □

Iwaniec [4] proved that there are infinitely many primes of the first form above, but we do not know if they satisfy the extra conditions on the bounds of s and t. The second condition above in particular implies that the upper bound $\left\lfloor \sqrt{4p/3} \right\rfloor$ can be achieved for infinitely many primes p if we *assume* a particular case of the Bunyakovsky conjecture [1]: there exist infinitely many primes p of the form $3s^2 - s - 1$ with $s > 0$.

Let $N \geqslant 2$ be an integer. We set $v := \lfloor \sqrt{N} \rfloor$, and set P_v be the convex hull of Farey_v^* in the hyperbolic plane $\mathscr{H}$. A starting point towards the above theorems is the observation that the projection

$$\pi \colon \mathscr{H} \to Y_0(N) := \mathscr{H}/\Gamma_0(N)$$

maps the interior of P_v injectively; see Lemma 3.3. This readily establishes an effective lower bound of $\lfloor \sqrt{N} \rfloor$ in the above theorem. The equality condition will follow from an asymptotic estimate of the number of ideal triangles in P_v.

The second key result we show is that under the extra hypothesis that $N = p$ or p^2 for a prime p, each connected component of the complement of $\pi(P_v)$ in $Y_0(N)$ is either an order–three cone or an ideal triangle; see Remark 3.11. Here, an *order–three cone* means the image of a triangle with angles $(0, 0, 2\pi/3)$ with the two rays at the angle $2\pi/3$ glued. For $N = p$ or p^2, we also have a geometric proof of the inequality

$$\left\lfloor \frac{N(1 + 1/p)}{3} \right\rfloor \geqslant \sum_{1 \leqslant i \leqslant \lfloor \sqrt{N} \rfloor} \phi(i),$$

by showing that the difference is precisely the number of such ideal triangles (constituting connected components). The equality holds iff N belongs to the list in Theorem 1.3 (1). As a consequence, this difference can be expressed in terms of the number of solutions to certain diophantine equations; see Remark 3.11. We also have an analogous description of $Y_0(N) \backslash \pi(P_v)$ for $N = pq$; see the same remark. Theorems 1.1 and 1.2 will be established in the process of these descriptions.

The very tight connection between the geometry and the arithmetic for the groups $\Gamma_0(N)$ established by the above results was somewhat surprising and hints at a deeper and more mysterious relation for congruence subgroups. For a general subgroup $\Lambda \leqslant \Gamma$ of index N, one cannot hope to obtain this type of bound, see Example 2.4 where $m(\Lambda)$ is exponential in N. The upper bound established above does depend on N being a prime or its square. For example, in $\Gamma_0(8)$ the upper bound above does not hold; see Example 2.7.

2. Preliminaries

In this section, we give a brief exposition on Kulkarni's construction of *special polygons* for congruence subgroups.

2.1. **Special polygons.** The group $\Gamma := \mathrm{PSL}(2, \mathbb{Z})$ acts faithfully on the upper half-plane $\mathscr{H}$ by orientation-preserving isometries, so that each element of Γ is identified with the corresponding linear fractional transformation. We let S be the rotation of angle π about $i = \sqrt{-1}$, and R be the rotation of angle $2\pi/3$ about $e^{\pi i/3}$.

We set $T(z) := R^{-1}S(z) = z+1$. The group $\Gamma = \langle S, R \mid S^2 = R^3 = 1\rangle$ leaves invariant the circularly ordered set

$$\hat{\mathbb{Q}} = \mathbb{Q} \cup \{\infty\} \subseteq \partial\mathscr{H} \approx S^1.$$

Unless specified otherwise, we always assume that an expression a/b for an element of $\hat{\mathbb{Q}}$ is in the *reduced form*, that is $\gcd(a,b) = 1$ and $b \geqslant 0$; such an expression is unique except for

$$\infty = 1/0 = (-1)/0 = -\infty.$$

For $x, y \in \partial\mathscr{H}$, the ordered pair (x,y) often means the oriented infinite geodesic in $\mathscr{H}$ moving from x to y.

A positively oriented triangle in the Γ–orbit of the geodesic triangle $\Delta_0 := (0,1,\infty)$ is called a *Farey triangle*. The *Farey tessellation* $\mathscr{T}$ is the 2–complex structure on $\mathscr{H} \cup \hat{\mathbb{Q}}$ having the Farey triangles as the 2–cells. Each triangle in the Γ–orbit of $\Delta_0^* := (0, e^{\pi i/3}, \infty)$ is called a *special triangle*. We let $\mathscr{T}^*$ denote the subdivision of $\mathscr{T}$ consisting of special triangles so that each Farey triangle is subdivided into three special triangles. The group Γ acts uniquely transitively on special triangles.

The pair of the endpoints of a geodesic line in the Farey tessellation is called a *Farey pair*. For a Farey pair (x,y) we define their *mediant* $x \oplus y$ as the unique vertex in $\hat{\mathbb{Q}}$ such that $(x, x\oplus y, y)$ is a (positively oriented, as usual) Farey triangle. Algebraically, $(a'/a, b'/b) \in \hat{\mathbb{Q}} \times \hat{\mathbb{Q}}$ is a Farey pair if $ab' - a'b = \pm 1$; if we further assume $a'/a, b'/b \in \mathbb{Q}$ and $a'/a < b'/b$, then we see that

$$\frac{a'}{a} \oplus \frac{b'}{b} = \frac{a'+b'}{a+b}.$$

For a geodesic polygon $P \subseteq \mathscr{H}$ we let $E(P)$ and $\mathrm{cusp}(P)$ denote the sets of the positively oriented sides and of the ideal vertices of P, respectively. For a side e, we let $\bar{e}$ denote its reversely oriented geodesic.

Definition 2.1. Let Λ be a subgroup of the classical modular group Γ.

(1) A *normalized ideal polygon* is a finite convex subcomplex P of $\mathscr{T}$ such that ∞ and 0 are vertices of P; in the case when $\Lambda = \Gamma_0(N)$ for some N, we further require that $(\infty, 0)$ is a positively oriented side of P;

(2) A *special polygon P^* for* Λ is a convex subcomplex of $\mathscr{T}^*$ which is a fundamental domain of Λ such that the convex hull of cusp P^* is normalized. In this case, each element in

$$\{g \in \Lambda \mid P^* \cap gP^* \text{ is a side of } P^*\}$$

is called a *side-pairing* corresponding to P^*.

(3) A normalized ideal polygon whose interior maps injectively under the quotient $\mathscr{H} \to \mathscr{H}/\Lambda$ is called a Λ–*injective ideal polygon*, or simply a Λ–*polygon*. Such a polygon is *maximal* if it is maximal under inclusion.

We observe that the convex hull of $\mathrm{cusp}(P^*)$ in part 2 is a maximal Λ–polygon. Note also that the set of side-pairings is closed under taking inverses. A remarkable discovery of Kulkarni in [6] is that a special polygon provides a geometric tool for the purely algebraic problem of efficiently finding an independent system for $\Lambda \leqslant \Gamma$.

Theorem 2.2 ([6]). *Every finite index subgroup Λ of Γ admits a special polygon; moreover, the corresponding set A of side-pairings contains an independent system B such that $A = B \cup B^{-1}$.*

A key idea in Kulkarni's construction is that every Λ–polygon P can be enlarged to some maximal Λ–polygon, and then to a special polygon for Λ; this follows from the fact that each side of such a P can be classified as below.

Lemma 2.3. *Let P be a Λ–polygon for some $\Lambda \leqslant \Gamma$. Suppose e is a side of P, and denote by Δ the unique Farey triangle that shares only the side e with P. Then exactly one of the following holds.*

(i) e is even*, meaning a unique $g_e \in \Lambda$ reverses the orientation of e.*
(ii) e is odd*, meaning a unique $g_e \in \Lambda$ rotates Δ by the angle $2\pi/3$;*
(iii) e is paired *to a different side f of P, meaning a unique $g_e \in \Lambda$ satisfies $\bar{f} = g_e(e)$;*
(iv) e is free*, meaning $P \cup \Delta$ is a Λ–polygon.*

In particular, we have a *labeling map*

$$\sigma\colon E(P) \to \{-4, -3, -2\} \cup \{0, 1, 2, 3, \ldots\}$$

so that the preimages of $-4, -3, -2$ are free, odd and even respectively, and the preimage of each positive integer is either empty or consists of two paired sides. We slightly abuse the terminology and say each element of

$$\{g_e \in \Lambda \mid e \text{ is a non-free side of } P\}$$

is a *side-pairing* for P. If P has no free sides, then P is maximal and included in some special polygon P^* with the same cusp set. Hence, the above set precisely consists of the side-pairings of P^* in the sense that was defined earlier. In the case when P is a maximal Λ–polygon with a labeling map σ, then the pair

$$(\operatorname{cusp} P, \sigma)$$

is often called a *Farey symbol* in the literature; see [6]. The set $\operatorname{cusp} P$ is a *generalized Farey sequence*, that is, a sequence of rational numbers whose adjacent terms are Farey pairs.

To construct a special polygon for a given group Λ, one starts with the Farey triangle $(0, 1, \infty)$ and continue adding a Farey triangle to a remaining free side to obtain a Λ–polygon. If Λ is of finite index, this process must terminate at a maximal Λ–polygon P. One then finally add a special triangle to each odd side of P, and obtain a special polygon as required in Theorem 2.2.

Conversely to Theorem 2.2, Kulkarni also noticed that if P is a finite area convex subcomplex of $\mathscr{T}$ with each side labeled by even, odd or a positive integer in such a way that each positive integer is assigned to either zero or two edges (called paired edges) then there uniquely exists a finite index subgroup $\Lambda_P \leqslant \Gamma$ such that the edges of P are classified as in the above lemma with respect to Λ_P, and such that Λ_P admits a special polygon P^* with $\operatorname{cusp}(P^*) = P$. In particular, Λ_P is generated by the side-pairings.

Example 2.4. Let us fix a positive integer N. We define a sequence

$$r_0 := 0 < r_2 < r_4 < \cdots < r_5 < r_3 < r_1 := 1$$

by the condition that $r_{i+2} = r_{i+1} \oplus r_i$ for $i = 0, \dots, N-2$. We note that each r_i coincides with $\text{Fibonacci}_i \,/\, \text{Fibonacci}_{i+1}$. We let $P(N)$ be the ideal polygon with the cusp set $\{r_i\}_{0 \leqslant i \leqslant N} \cup \{\infty\}$, and label the sides by the following rule.

$$\sigma(e) := \begin{cases} 1, & \text{if } e = (-\infty, 0) \text{ or } e = (1, \infty), \\ 2, & \text{if } e = (r_N, r_{N-1}) \text{ or } e = (r_N, r_{N-2}) \text{ up to orientation}, \\ -2, & \text{otherwise.} \end{cases}$$

Then there exists a finite index subgroup $\Lambda_{P(N)} \leqslant \Gamma$ generated by the side-pairings of $P(N)$. The group $\Lambda_{P(N)}$ admits a special polygon, which coincides with $P(N)$ since there are no odd sides. The vertices created converge to $1/\phi$ where ϕ is the golden ratio. The corresponding index satisfies

$$[\Gamma \colon \Lambda_{P(N)}]/3 = N,$$

which is the number of the Farey triangles in $P(N)$ after the final step. As there was essentially only one choice at each step of the construction once the initial ideal triangle $(0, 1, \infty)$ is chosen, this is the unique special polygon for $\Lambda_{P(N)}$ up to the action of Γ. We record for a later reference that the largest denominator in $\text{cusp}(P(N))$ is exponential in N.

2.2. **The congruence subgroup $\Gamma_0(N)$.** We are mostly concerned with the case $\Lambda = \Gamma_0(N)$. In the modular surface $Y_0(N) = \mathscr{H}/\Gamma_0(N)$, we denote by $v_\infty(N), v_2(N)$ and $v_3(N)$ the numbers of cusps, order–two cone points and order–three cone points respectively. It is well known that

$$[\Gamma : \Gamma_0(N)] = N \prod_{p \,|\, N} (1 + 1/p)\,.$$

Let $\mathscr{P}(N)$ denote the set of prime divisors of N. We have [8, Page 25] that

$$v_\infty(N) = \sum_{d \,|\, N} \phi\left(\gcd(d, N/d)\right),$$

and that

$$v_2(N) = \begin{cases} 0, & \text{if } 4 \mid N \text{ or if } \mathscr{P}(N) \cap (4\mathbb{Z} - 1) \neq \varnothing, \\ 2^k, & \text{otherwise, with } k := \#\left(\mathscr{P}(N) \cap (4\mathbb{Z} + 1)\right), \end{cases}$$

$$v_3(N) = \begin{cases} 0, & \text{if } 9 \mid N \text{ or if } \mathscr{P}(N) \cap (3\mathbb{Z} - 1) \neq \varnothing, \\ 2^k, & \text{otherwise, with } k := \#\left(\mathscr{P}(N) \cap (3\mathbb{Z} + 1)\right). \end{cases}$$

The genus g can be computed using the Riemann–Hurwicz formula. Namely,

$$\frac{\pi}{3}[\Gamma : \Gamma_0(N)] = \text{Area}(\mathscr{H}/\Gamma_0(N)) = -2\pi \left(2 - 2g - v_\infty(N) - \frac{2v_3(N)}{3} - \frac{v_2(N)}{2}\right).$$

The number of each free factor in the following decomposition is computed:

$$\Gamma_0(N) \cong \left(\mathop{\ast}_{i=1}^{v_2(N)} \mathbb{Z}/2\mathbb{Z}\right) * \left(\mathop{\ast}_{i=1}^{v_3(N)} \mathbb{Z}/3\mathbb{Z}\right) * \left(\mathop{\ast}_{i=1}^{2g+v_\infty(N)-1} \mathbb{Z}\right).$$

One particular advantage of the approach in Theorem 2.2 is that matrix representatives of an independent system can be read off directly during the process

since the choice of the element g_e is straightforward from the label and the endpoints. Let us record one consequence that will be relevant to our purpose. Recall our convention that rational numbers are assumed to be written in reduced form.

Lemma 2.5 ([6, 2])**.** *Let P be a $\Gamma_0(N)$–polygon, and let $e = (a'/a, b'/b)$ be a side of P not incident at ∞. Then the following hold.*

(1) The side e is even iff $N \mid a^2 + b^2$;
(2) The side e is odd iff $N \mid a^2 + ab + b^2$;
(3) The side e is paired to another side $f = (c'/c, d'/d)$ iff $N \mid ac + bd$.

Suppose P is a $\Gamma_0(N)$–polygon, whose cusp set is written as

$$x_{-1} = -\infty < x_0 = 0 < x_1 < \cdots < x_\ell = 1 < \infty = x_{-1}.$$

In particular, the two sides $(-\infty, 0)$ and $(1, \infty)$ are paired. The number ℓ of ideal triangles in a special polygon P^* for $\Gamma_0(N)$ is given by

$$u(N) := \frac{\mathrm{Area}(P^*) - \pi v_3(N)/3}{\pi} = \frac{[\Gamma\colon \Gamma_0(N)] - v_3(N)}{3} = \#\,\mathrm{cusp}(P^*) - 2.$$

In particular, if N is a nontrivial power of a prime p then $u(N) = \lfloor N(1 + 1/p)/3 \rfloor$. The labeling map σ of P is often succinctly encoded as a tuple

$$(\sigma(-\infty, x_0 = 0) := 1, \sigma(x_0, x_1), \ldots, \sigma(x_\ell = 1, \infty) := 1).$$

In general, the *denominator sequence* of a sequence $A = \{b'_i/b_i\}$ in $\hat{\mathbb{Q}}$ is defined as

$$d(A) := dA = \{b_i\}.$$

Let us now assume the denominator sequence of cusp P is written as

$$d(\mathrm{cusp}\, P) := \{a_{-1} = 0, a_0 = 1, a_1, \ldots, a_\ell = 1\}.$$

We call each consecutive pair (a_i, a_{i+1}) a *side* of $d(\mathrm{cusp}\, P)$, with indices taken cyclically. Such a side (a_i, a_{i+1}) inherits its labeling from the side $(x_i, x_{i+1}) \in E(P)$.

A $\Gamma_0(N)$–polygon P is uniquely determined by the denominator sequence of $\mathrm{cusp}(P)$; this is due to the following general observation.

Lemma 2.6. *If (a, b) is a pair of positive co-prime integers, then there uniquely exist nonnegative integers a', b' such that a'/a and b'/b are in reduced forms comprising a Farey pair, and such that*

$$0 \leqslant a'/a < b'/b \leqslant 1.$$

Proof. The equation $ax - by = 1$ admits a unique solution (x, y) satisfying $0 \leqslant y < a$ and $1 \leqslant x \leqslant b$. This is equivalent to the condition that y/a and x/b are a Farey pair in $[0, 1]$, as required. □

From the two preceding lemmas, we see that it suffices to work only with the denominator sequence when finding and enlarging a $\Gamma_0(N)$–polygon P.

Example 2.7. The denominator sequence

$$d(\mathrm{cusp}\, P) = \{0, 1, 4, 3, 2, 1\}$$

uniquely determines a normalized ideal polygon $P \leqslant \mathscr{T}$ as

$$\mathrm{cusp}(P) = \{-\infty, 0, 1/4, 1/3, 1/2, 1\}.$$

In the case when $N = 8$, it is easy to deduce from Lemma 2.5 and from the denominator sequence that the labeling map of P is given as

$$\sigma = (1, 2, 2, 3, 3, 1).$$

See Figure 1. Since there are no odd or free sides, we see P is a special polygon for $\Gamma_0(8)$. There exists a different special polygon for $\Gamma_0(8)$. Namely, one can verify that $d(\text{cusp } P') = \{0, 1, 2, 3, 4, 1\}$ and $\sigma' = (1, 3, 3, 2, 2, 1)$ also determine a special polygon P' for the same group.

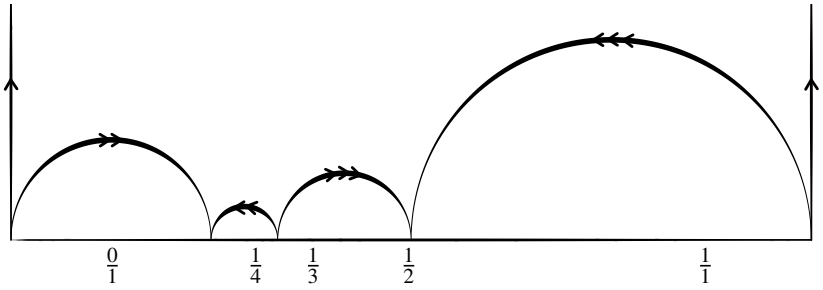

FIGURE 1. One of the two possible special polygons for $\Gamma_0(8)$. The denominator sequence is $\{0, 1, 4, 3, 2, 1\}$.

Example 2.8. Let us work out all possible denominator sequences and the labelings of maximal ideal polygons P for $\Gamma_0(N)$ with $N = 2, 3, 5$ and 7.

- $N = 2$: $d(\text{cusp } P) = \{0, 1, 1\}$, $\sigma = (1, -2, 1)$.
- $N = 3$: $d(\text{cusp } P) = \{0, 1, 1\}$, $\sigma = (1, -3, 1)$.
- $N = 5$: $d(\text{cusp } P) = \{0, 1, 2, 1\}$, $\sigma = (1, -2, -2, 1)$.
- $N = 7$: $d(\text{cusp } P) = \{0, 1, 2, 1\}$, $\sigma = (1, -3, -3, 1)$.

Note that $N = 2, 3$ and $N = 5, 7$ share the same cusps with different labelings.

3. Bounds on the denominators of cusps

As briefly discussed in Section 1, a key step in our proof for Theorems 1.1 and 1.2 is to establish effective lower and upper bounds for the following quantity:

$$\begin{aligned} m(\Gamma_0(N)) &:= \min \{\max d(\text{cusp } P) \mid P \text{ is a maximal } \Gamma_0(N)\text{–polygon}\} \\ &= \min \{\max d(\text{cusp } P^*) \mid P^* \text{ is a special polygon for } \Gamma_0(N)\} \end{aligned}$$

A special polygon P^* for $\Gamma_0(N)$ is *optimal* if the maximum of $d(\text{cusp}(P^*))$ coincides with $m(\Gamma_0(N))$.

For each N there exist only finitely many possible maximal (normalized, by definition) ideal polygons for $\Gamma_0(N)$. Hence, there exists a finite time algorithm to determine the value $m(\Gamma_0(N))$. Let us record some of such values found by a computer search in Table 1, for a later use.

N	6	10	14	23
$m(\Gamma_0(N))$	3	5	7	5

TABLE 1. The values of $m(\Gamma_0(N))$.

In the first subsection, we prove part 1 of Theorem 1.3; Parts 2 and 3 of the same theorem are proved in the second subsection, where Theorems 1.1 and 1.2 are also deduced. *Throughout this section, we let $N \geqslant 2$ be an integer, and set*

$$v := \lfloor \sqrt{N} \rfloor.$$

We let P_v denote the convex hull of Farey_v^*.

3.1. **Establishing the lower bound.** Recall we have denoted by $u(N)$ the number of ideal triangles in a maximal $\Gamma_0(N)$–polygon. The *totient summatory function* is defined as

$$\Phi(n) := \sum_{i=1}^{n} \phi(i) = \#\,\mathrm{Farey}_n^* - 2.$$

We have the following preliminary bounds for $m(\Gamma_0(N))$.

Lemma 3.1. *For all integer $N \geqslant 2$ we have that*

$$\min \Phi^{-1}[u(N), \infty) \leqslant m(\Gamma_0(N)) \leqslant ((1 + \sqrt{5})/2)^{u(N)}.$$

Furthermore, we have that

$$\liminf_{N\to\infty} \frac{m(\Gamma_0(N))}{[\Gamma\colon \Gamma_0(N)]^{1/2}} \geqslant \frac{\pi}{3}.$$

Proof. Fix an integer $N \geqslant 2$ and set $m := m(\Gamma_0(N))$. Since there exists a maximal $\Gamma_0(N)$–polygon Q whose cusp set is contained in Farey_m^*, we have

$$u(N) = \#\,\mathrm{cusp}(Q) - 2 \leqslant \#\,\mathrm{Farey}_m^* - 2 = \Phi(m).$$

This gives the first inequality. For the second one, it suffices to note that each cusp of Q belongs to a Farey triangle that can be joined to the Farey triangle $(0, 1, \infty)$ by a sequence of adjacent Farey triangles, and that such a sequence has at most $u(N)$ triangles; see also Example 2.4.

To see the last part of the lemma, we note the classical fact [10] that

$$\Phi(m) = 3m^2/\pi^2 + O(m(\log m)^{2/3}(\log\log m)^{4/3}).$$

For $N \geqslant 4$, we claim that the (crude) bound

$$v_3(N) \leqslant \sqrt{N}$$

holds. Indeed, suppose $N = p_1^{e_1} \cdots p_k^{e_k}$ is the prime factorization with

$$p_1 < \cdots < p_k.$$

We only consider the case that $k \geqslant 2$ and $p_i \in \{3\} \cup (3\mathbb{Z} + 1)$ for all i, since the claim is trivial otherwise. We observe

$$p_1 \cdots p_k \geqslant 3 \cdot \prod_{i=1}^{k-1} (6i + 1) \geqslant 4^k.$$

It follows that $v_3(N) = 2^k \leqslant \sqrt{p_1 \cdots p_k} \leqslant \sqrt{N}$.

Assuming N is sufficiently large, we have that

$$[\Gamma\colon \Gamma_0(N)] - \sqrt{N} \leqslant 3u(N) \leqslant 3\Phi(m) \leqslant 9m^2/\pi^2 + 3m\log m.$$

Since $[\Gamma\colon \Gamma_0(N)] > N$, we obtain the latter part of the conclusion. □

The following elementary observation will be useful for us.

Lemma 3.2. *If a, b, x, y are integers in $(0, v]$ such that $x + y \leqslant v$ and such that $\gcd(a, b) = 1$, then we have*

$$N \nmid ax + by, \quad N \nmid x^2 + y^2, \quad N \nmid x^2 + xy + y^2.$$

Proof. Since a and b are co-prime, we have $\min(a, b) < v$. It follows that

$$2 \leqslant ax + by < v(x + y) \leqslant N.$$

We also have that $x^2 + y^2 < x^2 + xy + y^2 < (x + y)^2 \leqslant N$, implying the conclusion. □

A parabolic element in Γ is a *translation* iff it fixes ∞.

Lemma 3.3. *We have that P_v is a $\Gamma_0(N)$–polygon; furthermore, the $(2, 1)$ component of each non-translation side-pairing for P_v is N.*

Proof. Let $(\bar{a}, \bar{b})$ and $(\bar{x}, \bar{y})$ be Farey pairs from Farey_v such that

$$\bar{a} < \bar{b}, \quad \bar{x} < \bar{y},$$

and such that $\bar{a}$ and $\bar{b}$ are non-adjacent in Farey_v. Let us denote by a, b, x, y the corresponding denominators, so that $a + b \leqslant v$. Lemma 3.2 implies that $N \nmid ax + by$. Using Lemma 2.5, we see that the geodesics $(\bar{a}, \bar{b})$ and $(\bar{x}, \bar{y})$ are never identified, even up to orientations, under the action of $\Gamma_0(N)$; this implies that P_v is a $\Gamma_0(N)$–polygon.

Suppose now that C is the $(2, 1)$ component of a side-pairing g that does not fix ∞. Let (a, b) be a non-free side e of $d\,\mathrm{Farey}_v^*$ so that $g_e = g$. Note that $(a, b) \neq (0, 1)$ or $(1, 0)$. If e is paired or even, then we see from Lemma 2.5 that some (possibly the same) side (x, y) of $d\,\mathrm{Farey}_v^*$ satisfies

$$N \mid C = ax + by.$$

Since $a, b, x, y \leqslant \sqrt{N}$ and $\gcd(a, b) = 1$, we have that

$$0 < C = ax + by \leqslant (a + b)\max(x, y) < 2N,$$

and that $C = N$. If (a, b) is odd, we also see from Lemma 2.5 that

$$N \mid C = a^2 + ab + b^2 < 3N.$$

Using a simple parity argument, we conclude that $C = N$. □

Let us record an algebraic consequence of our discussion so far.

Lemma 3.4. *For all integer $N \geqslant 2$, we have that*

$$\Phi(N) \leqslant \frac{[\Gamma : \Gamma_0(N^2)] - \nu_3(N^2)}{3}.$$

When N is a power of a prime p, we also have that

$$\Phi(\lfloor \sqrt{N} \rfloor) \leqslant \left\lfloor \frac{N(1 + 1/p)}{3} \right\rfloor.$$

Proof. Lemma 3.3 implies

$$\Phi(\lfloor \sqrt{N} \rfloor) = \#\,\mathrm{cusp}(P_v) - 2 \leqslant u(N),$$

from which the conclusions are immediate. □

Let us also note a (non-asymptotic) bound

$$\Phi(m) \leqslant 3m^2/\pi^2 + 2m \log m$$

for all $m \geqslant 2$. We can now establish the lower bound for $m(\Gamma_0(N))$.

Proof (of part 1 in Theorem 1.3). From Lemmas 3.4 and 3.1, we see that $\Phi(v) \leqslant u(N)$, and hence, that $m(\Gamma_0(N)) \geqslant v$.

The equality $m(\Gamma_0(N)) = v$ holds iff the polygon P_v is maximal, which is equivalent to that $u(N) = \Phi(v)$. As in the proof of Lemma 3.1, this equality would imply

$$\frac{N - \sqrt{N}}{3} < u(N) = \Phi\left(\lfloor \sqrt{N} \rfloor\right) < \frac{3}{\pi^2} N + \sqrt{N} \log N.$$

Note that the following function is strictly positive for all $x > 10^6$:

$$f(x) := \frac{x - \sqrt{x}}{3} - \frac{3}{\pi^2} x - \sqrt{x} \cdot \log x.$$

On the other hand, it is a routine computer verification to see that the equation $u(N) = \Phi(v)$ holds for an integer $N < 10^6$ iff N is in the given list of Theorem 1.3. This completes the proof. □

Remark 3.5. From the proof, it is clear that an integer $N \geqslant 2$ belongs to the list in part 1 of Theorem 1.3, if and only if for each positive coprime pair (a, b) of integers satisfying $a, b \leqslant \sqrt{N}$, there exists a positive co-prime pair (x, y) such that

$$ax + by = N,$$

and such that $x, y \leqslant \sqrt{N}$.

3.2. **Farey triple.** A crucial idea for the remaining parts of Theorem 1.3 is that, under given hypotheses we can group many (sometimes all) of the free sides of $d\,\mathrm{Farey}_v^*$ into *Farey N–triples*, each of which consists of three sides with certain properties that control the largest denominators in the resulting special polygons. Recall the notations S, R, T for the matrices given in the introduction.

Definition 3.6. A sequence of distinct co-prime pairs of positive integers

$$\{(a_i, b_i) \colon i = 0, 1, 2\}$$

is called an *N–Farey triple* if the following holds for $i = 0, 1, 2$ with the indices taken cyclically:

$$\begin{pmatrix} a_{i+1} & b_{i+1} \end{pmatrix} T \begin{pmatrix} a_i \\ b_i \end{pmatrix} = N.$$

Let $\{(a_i, b_i) \mid i = 0, 1, 2\}$ be an N–Farey triple, and pick the unique Farey pair $(\bar{a}_i, \bar{b}_i)$ corresponding to each (a_i, b_i); see Lemma 2.6. By Lemma 2.5 and by the unique transitivity of the action of Γ on special triangles, we have a unique $\gamma_i \in \Gamma_0(N)$ such that

$$\left(\bar{a}_i, \bar{a}_i \oplus \bar{b}_i, \bar{b}_i\right) \xrightarrow{\gamma_i} \left(\bar{a}_{i+1} \oplus \bar{b}_{i+1}, \bar{b}_{i+1}, \bar{a}_{i+1}\right).$$

The identifications of these triangles are illustrated in Figure 2 by single, double or triple arrows.

The following lemma explains the connection between the (algebraically defined) N–Farey triples and the (geometrically defined) sides of $d\,\mathrm{Farey}_v^*$. It further implies that the first two pairs in an N–Farey triple determine the remaining pair.

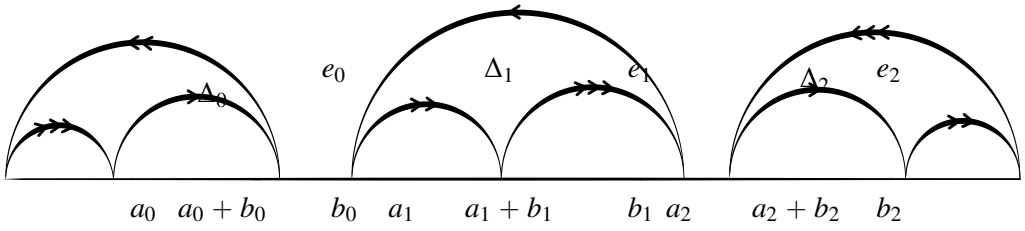

FIGURE 2. An N–Farey triple realized as denominators of Farey pairs. Single, double or triple arrows illustrate how these sides are identified by the action of $\Gamma_0(N)$.

Lemma 3.7. *Suppose we have a sequence of co-prime pairs of positive integers*

$$Z = \{(a_i, b_i) \colon i = 0, 1, 2\}.$$

(1) If Z is an N–Farey triple, then we have that $a_i + b_i = b_{i+1} + a_{i+2}$ for each i, and that $\min_i(a_i + b_i) < \sqrt{4N/3}$. If we further assume that $\min_i(a_i + b_i) > \sqrt{N}$, then we have that Z consists of free sides in $d\,\mathrm{Farey}_v^*$ and that $\max(a_i, b_i) < \sqrt{N}$ for all i.

(2) Conversely, if we have that $a_1(a_0 + b_0) + b_1 b_0 = N$ and that

$$(a_2, b_2) = (a_0 + b_0 - b_1, -a_0 + a_1 + b_1),$$

then either $Z = \{(a_0, b_0)\}$ or Z is an N–Farey triple.

(3) If Z consists of sides in $d\,\mathrm{Farey}_v^*$, and if

$$a_{i+1}(a_i + b_i) + b_{i+1} b_i = N$$

for $i = 0, 1$, then either $Z = \{(a_0, b_0)\}$ or Z is an N–Farey triple.

Proof. 1. We note that

$$\begin{pmatrix} N \\ N \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} a_1 & b_1 \end{pmatrix} T^t \\ \begin{pmatrix} a_0 & b_0 \end{pmatrix} T \end{pmatrix} \begin{pmatrix} a_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 + b_1 & b_1 \\ a_0 & a_0 + b_0 \end{pmatrix} \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}.$$

The matrix in the right-most term has the determinant N and hence,

$$\begin{pmatrix} a_2 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_0 + b_0 & -b_1 \\ -a_0 & a_1 + b_1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} a_0 + b_0 - b_1 \\ -a_0 + a_1 + b_1 \end{pmatrix}.$$

This proves the claimed equality. We also note for each i that

$$a_i a_{i+1} + (a_i + b_i) b_{i+1} = N \Leftrightarrow (a_i + b_i)(a_{i+1} + b_{i+1}) - N = b_i a_{i+1}.$$

Setting $A = \min_i(a_i + b_i)$, we obtain that

$$\left(1 - \frac{N}{A^2}\right)^3 \leqslant \prod_i \left(1 - \frac{N}{(a_i + b_i)(a_{i+1} + b_{i+1})}\right) = \prod_i \frac{a_i b_i}{(a_i + b_i)^2} < \frac{1}{64}.$$

This implies that $A < \sqrt{4N/3}$, as required.

For the latter conclusion, let us assume $a_i + b_i > \sqrt{N}$ for each i. We note

$$a_{i+1}\sqrt{N} < a_{i+1}(a_i + b_i) < a_{i+1}(a_i + b_i) + b_{i+1} b_i = N.$$

By similar consideration for b_i, we have that $b_i < \sqrt{N}$ for all i. Since $\gcd(a_i, b_i) = 1$, we see from Lemma 2.6 that $d\,\mathrm{Farey}_v^*$ has (a_i, b_i) as a side.

Assume for contradiction that (a_0, b_0) is not a free side. If (a_0, b_0) is a paired side or an even side, then there exists a side (a, b) in $d\,\mathrm{Farey}_v^*$ such that $a_0a + b_0b = N$. In the case when $b > a$, we deduce from

$$a(a_0 + b_0) + (b - a)b_0 = N = a_1(a_0 + b_0) + b_1b_0$$

and from $0 < a_1, b_1 \leqslant v < a_0 + b_0$ that

$$(a_1, b_1) = (a, b - a).$$

This contradicts $a_1 + b_1 > v$. Similarly, if $b < a$, we would see from

$$N = b(a_0 + b_0) + (a - b)a_0 = b_2(a_0 + b_0) + a_2a_0$$

that $(a_2, b_2) = (a - b, b)$, which again is a contradiction.

If (a_0, b_0) is an odd side, then we would have that

$$N = a_0(a_0 + b_0) + b_0^2 = a_1(a_0 + b_0) + b_1b_0,$$

and that $(a_1, b_1) = (a_0, b_0)$. This is again a contradiction, and we conclude that each (a_i, b_i) is a free side.

2. Simply reading backward the former half of the proof of part 1, we verify the equation in Definition 3.6 holds for $i = 0, 1, 2$. It is easy to see from the hypothesis that either Z is a singleton or Z consists of three distinct pairs.

3. We have two solutions $(a_0 + b_0, b_0)$ and $(a_2, a_2 + b_2)$ of the equation

$$a_1x + b_1y = N.$$

Using the hypothesis that (a_i, b_i) is a side, we see

$$(a_2, a_2 + b_2) = (a_0 + b_0 - b_1, b_0 + a_1).$$

Then we compute

$$(a_2 + b_2)a_0 + b_2b_0 = (b_0 + a_1)a_0 + (a_1 + b_1 - a_0)b_0 = N.$$

The conclusion now follows from part (2). □

We say a set $A \subseteq \mathbb{Z}^2$ is said to be *primitive* if each vector $(x, y) \in A$ satisfies that $\gcd(x, y) = 1$. Let us introduce the notation

$$S(a, b; N) := \{(x, y) \in \mathbb{Z}^2 \mid ax + by = N\}.$$

Lemma 3.8. *If (a, b) is a free side of $d\,\mathrm{Farey}_v^*$, and if the set*

$$S(a, b; N) \cap ((v - b, v] \times \mathbb{Z}_+ \cup \mathbb{Z}_+ \times (v - a, v])$$

is primitive, then (a, b) belongs to an N–Farey triple, which consists of free sides in $d\,\mathrm{Farey}_v^*$.

Proof. Let us establish a series of claims for the proof.

Claim 1. *If $(x, y) \in S(a, b; N) \cap (0, v]^2$, then $x + y > v$ and $\gcd(x, y) \neq 1$.*

Let (x, y) be as in the hypothesis. We note from Lemma 3.2 that $x + y > v$. If $\gcd(x, y) = 1$ then (x, y) would be a side of $d\,\mathrm{Farey}_v^*$, which contradicts the assumption that (a, b) is free; see Lemma 2.5. The claim follows.

Claim 2. *We have that $\max(a, b) < \sqrt{N}$.*

Suppose $b < a \leqslant v$. Since $\gcd(a-b,a) = 1$, we see from Claim 1 that

$$(a-b,a) \notin S(a,b;N).$$

This proves $a^2 \neq N$. Similarly, we have $b^2 \neq N$, as claimed.

Let us now consider the unique pair $(x_2,y_2) \in S(a,b;N)$ such that

$$0 \leqslant v-b < x_2 \leqslant v.$$

Since $x_2+b, a+b \in [v+1,\infty) \subseteq (\sqrt{N},\infty)$, we have

$$0 < N-av \leqslant N-ax_2 = by_2 < \sqrt{N}(x_2+b) - ax_2 < b(x_2+\sqrt{N}).$$

In particular, we have $0 < y_2 < x_2 + \sqrt{N}$. We see from the hypothesis that $\gcd(x_2,y_2) = 1$. Claim 1 implies $y_2 > v \geqslant x_2$.

Similarly, the unique $(x_0,y_0) \in S(a,b;N)$ satisfying

$$0 \leqslant v-a < y_0 \leqslant v,$$

will have the properties that $\gcd(x_0,y_0) = 1$ and that

$$y_0 \leqslant v < x_0 < y_0 + \sqrt{N}.$$

Then $(a_1,b_1) := (a,b)$, $(a_2,b_2) := (x_2, y_2 - x_2)$ and $(a_0,b_0) := (x_0 - y_0, y_0)$ satisfy the equation in Definition 3.6 with $i = 0,1$. Furthermore, $Z := \{(a_i,b_i)\}_{i=0,1,2}$ is not a singleton since (a,b) is not odd. Lemma 3.7 (3) implies that Z is an N–Farey triple consisting of free sides. □

We emphasize, we have not assumed that N is a prime or its square until this point.

Lemma 3.9. *If $N = p$ or $N = p^2$ for a prime p, and if (a,b) is a free side in $d\,\mathrm{Farey}_v^*$, then the solution set $S(a,b;N) \cap \mathbb{Z}_+^2$ is primitive.*

Proof. Let $(x,y) \in \mathbb{Z}_+^2$ satisfy $ax + by = N$. We have that

$$\gcd(x,y) \leqslant \min(x,y) \leqslant N/(a+b) < \sqrt{N}.$$

This implies that $\gcd(x,y) = 1$. □

We note the following simple fact on Möbius transformations.

Lemma 3.10. *Let us consider a matrix*

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$$

with $c > 0$. If there exist some rational numbers $0 \leqslant x < y \leqslant 1$ satisfying

$$0 \leqslant g(y) < g(x) \leqslant 1,$$

then we have that $|\operatorname{tr} g| \leqslant c-2$ and that $\|g\|_F < 2c-1$.

Proof. Using the equality $ad - bc = 1$ and the sequence of inequalities

$$0 \leqslant g(y) \leqslant g(1) < g(\infty) < g(0) \leqslant g(x) \leqslant 1,$$

we can establish

$$-c < d \leqslant b < 0 < -b \leqslant a < c.$$

The conclusion is then immediate. □

Proof (of part 2 of Theorem 1.3 and Theorem 1.1). Let $N = p$ or $N = p^2$ for a prime p. It is clear from the preceding lemmas that every free side of $d\,\mathrm{Farey}_v^*$ belongs to a unique N–Farey triple. We can hence partition the free sides of $d\,\mathrm{Farey}_v^*$ into a collection of N–Farey triples

$$\bigsqcup_{i=1,\dots,k} \{(a_j^i, b_j^i) \colon j = 0, 1, 2\}$$

such that for all i we have

$$a_0^i + b_0^i = \min_j (a_j^i + b_j^i) \in \left(\sqrt{N},\, \sqrt{4N/3} \right).$$

There exists a Farey pair $(\bar{a}^i, \bar{b}^i)$ corresponding to the free side (a_0^i, b_0^i). Set

$$c_i := \bar{a}^i \oplus \bar{b}^i \in \mathrm{Farey}^*_{\lfloor \sqrt{4N/3} \rfloor}.$$

We have proved above that the newly added sides $(\bar{a}^i, c_i)$ and $(c_i, \bar{b}^i)$ are paired to two of the free sides in Farey_v^*. In this way, we obtain all the side-pairings for

$$P' := \mathrm{hull}(\mathrm{Farey}_v^* \sqcup \{c_i\}_i),$$

which is a maximal $\Gamma_0(N)$–polygon. Part 2 of Theorem 1.3 follows from

$$m(\Gamma_0(N)) \leqslant \max_{i=1,\dots,k} \min_{j=0,1,2} (a_j^i + b_j^i) \leqslant \left\lfloor \sqrt{4N/3} \right\rfloor.$$

Combining the definition of a Farey N–triple with Lemma 3.3, we see that the $(2,1)$ component of each side-pairing for P' is N, with the sole exception of the translation $(x, y) \mapsto (x \pm 1, y)$. The desired bounds in Theorem 1.1 now follow from Lemma 3.10. □

Proof (of part 3 of Theorem 1.3 and Theorem 1.2). Let us assume $\sqrt{p} < \sqrt{q} < \sqrt{p} + \sqrt{2}$. If $p = 2$, then we have that $N \in \{6, 10, 14\}$, in which case part 3 follows by Table 1. So, we may assume that p and q are odd primes. We set $2k = q - p$.

The given bound on $\sqrt{q} - \sqrt{p}$ implies $v = p + k - 1$. We note that each element in the set

$$A := \{(k, p), (p, k)\} \cup \{(i, q - i) \mid k < i < p + k\}$$

is a free side of P_v. Furthermore, if (a, b) is a free side of P_v not belonging to A, then it is not hard to see that the set

$$S(a, b; N) \cap ((v - b, v] \times \mathbb{Z}_+ \cup \mathbb{Z}_+ \times (v - a, v])$$

is primitive. Applying Lemma 3.8 and suitably splitting the sides in A, we obtain another $\Gamma_0(N)$–polygon P' each of whose sides is exactly one of the following types:

(i) a non-free side of P_v;
(ii) a free side (x, y) of P_v such that the pair (x, y) belongs to an N–Farey triple;
(iii) a side in the set $A' := \{(p + k, p), (p, k)\}$;
(iv) a side in the set $A'' := \{(i, q), (q, p - i) \mid k \leqslant i \leqslant p - k - 1\}$;
(v) a side in the set $A''' := \{(p - k + i, q), (q, p + k - i) \mid 0 \leqslant i \leqslant 2k - 1\}$.

The sides in A', A'' and A''' can be paired among themselves respectively; concretely, we have

$$\begin{aligned}(p+k,p)\cdot(p,k) &= N.\\ (i,q)\cdot(q,p-i) &= N.\\ (p-k+i,q)\cdot(q,p+k-i) &= 2N.\end{aligned}$$

In particular, P' is a maximal $\Gamma_0(N)$–polygon. Using a similar argument to the case of $N = p$ or p^2, we obtain the desired upper bound. For the lower bound, we note that the denominator of a rational number that is in the orbit of $1/q$ by the action of $\Gamma_0(N)$ must be a multiple of q [8]. □

Remark 3.11. (1) For a positive integer N, let us denote by $k(N)$ the number of the N–Farey triples among the (necessarily free) sides in $d\,\mathrm{Farey}_v^*$. Algebraically, this number coincides with the number of ordered pairs (a_0, b_0) of co-prime positive integers such that for some choices of positive integers a_1, b_1, a_2 and b_2, all of the following hold:

(i) For each i with the indices taken cyclically, we have:

$$\begin{pmatrix} a_{i+1} & b_{i+1}\end{pmatrix} T \begin{pmatrix} a_i & b_i\end{pmatrix}^t = N;$$

(ii) (a_i, b_i) is a co-prime pair for each i;

(iii) We have

$$\sqrt{N} < a_0 + b_0 = \min_i(a_i + b_i) < a_1 + b_1.$$

Condition (iii) guarantees that one Farey N–triple is not counted more than once. Note we allow $a_0 + b_0 = a_2 + b_2$.

(2) Let N be a positive integer, and let π denote the quotient map from $\mathscr{H}$ onto the orbifold $\mathscr{H}/\Gamma_0(N)$. We define P'_v as the union of $P_v = \mathrm{hull}(\mathrm{Farey}_v^*)$ with one special triangle glued on each odd side. In the case when N is a prime or its square, our proof implies the existence of one-to-one correspondence between the set of the connected components in $(\mathscr{H}/\Gamma_0(N))\setminus\pi(P'_v)$ and the set of Farey N–triples $\{(a_i, b_i)\}_i$ satisfying $a_i + b_i > \sqrt{N}$ modulo the cyclic $\mathbb{Z}/3\mathbb{Z}$–symmetry. As a result, we have an estimate of the Euler totient summatory function

$$\sum_{i=1}^{\lfloor\sqrt{N}\rfloor} \phi(i) = u(N) - k(N) = \left\lfloor \frac{N(1+1/p)}{3} \right\rfloor - k(N).$$

(3) In the case when $N = pq$ as in Theorem 1.2, each connected component of $(\mathscr{H}/\Gamma_0(N))\setminus\pi(P'_v)$ is either a Farey triangle corresponding to some Farey N–triple, or an ideal polygon with a puncture inside; see Figure 3 for the description of this ideal polygon when $q = p + 2$. Such an ideal polygon has $p + 1$ Farey triangles. Hence, we have an estimate

$$\sum_{i=1}^{\lfloor\sqrt{N}\rfloor} \phi(i) + k(N) + p + 1 = u(N) = \frac{(p+1)(q+1) - v_3(pq)}{3}.$$

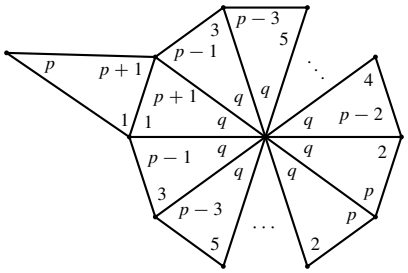

FIGURE 3. A complementary region of hull (Farey_v^*) in $\mathscr{H}/\Gamma_0(pq)$ with $q = p + 2$. Three corners of each Farey triangle are labeled with the denominators of corresponding cusp points.

4. Optimality of the upper bound

We will prove part 4 of Theorem 1.3 in this section, hence establishing the optimality of the coefficient $\sqrt{4/3}$ in the theorem. The crucial notion to consider is the following.

Definition 4.1. Let us say an integer N is *cashew* iff there exist nonnegative integers s, t, a, b such that

$$\begin{cases} N = sa + tb, \\ s + t > a = \lfloor \sqrt{4N/3} \rfloor > t \geqslant b \geqslant a - s \geqslant 0. \end{cases}$$

The following is the list of cashew integers up to 36:

$$4, 5, 8, 9, 10, 11, 14, 15, 16, 17, 18, 21, 22, 23, 24, 25, 26, 30, 32, 33, 34, 35, 36.$$

We will utilize an alternative formulation of cashew integers as below.

Lemma 4.2. *An integer N is cashew iff $\sqrt{N} \in \{2, 3, 4, 5, 6\}$, or there is some N–Farey triple $\{(a_i, b_i)\}$ such that*

$$\min_i (a_i + b_i) = \left\lfloor \sqrt{4N/3} \right\rfloor.$$

Proof. It is a routine check that $\sqrt{N} \in \{2, 3, 4, 5, 6\}$ if and only if N is cashew with b being zero in Definition 4.1. Let us now assume N is not one of these five numbers.

For the backward direction, let us pick a corresponding N–Farey triple $\{(a_i, b_i)\}$. We may assume that

$$v := \left\lfloor \sqrt{N} \right\rfloor \leqslant \left\lfloor \sqrt{4N/3} \right\rfloor = a_0 + b_0 = \min_i (a_i + b_i) < \max_i (a_i + b_i) = a_1 + b_1,$$

after applying a cyclic permutation to $\{(a_i, b_i)\}$, or changing the roles of a_i's and b_i's if necessary; see Lemma 3.7. Setting $(s, t, a, b) := (a_1, b_1, a_0 + b_0, b_0)$, we have the following.

- $sa + tb = a_1(a_0 + b_0) + b_1 b_0 = N$;
- $s + t - a = a_1 + b_1 - a_0 - b_0 > 0$;
- $a - t = a_0 + b_0 - b_1 = a_2 > 0$;
- $t - b = b_1 - b_0 = (a_1 + b_1) - (b_0 + a_1) = (a_1 + b_1) - (a_2 + b_2) \geqslant 0$;

- $b-(a-s)=b_0-(a_0+b_0)+a_1=(a_2+b_2)-(a_0+b_0)\geqslant 0$;
- $a-s=a_0+b_0-a_1\geqslant \left\lfloor \sqrt{4N/3}\right\rfloor - v\geqslant 0$;

This verifies that N is cashew.

For the forward direction, suppose the quadruple (s,t,a,b) certifies that N is cashew as in Definition 4.1, with $b>0$. We can uniquely choose a triple $\{(a_i,b_i)\}$ of positive pairs such that the following hold:

$$(s,t,a,b)=(a_1,b_1,a_0+b_0,b_0),$$
$$(a_2,b_2)=(a-t,s+t-(a-b))=(a_0+b_0-b_1,-a_0+a_1+b_1).$$

By Lemma 3.7 (2), the triple is an N–Farey triple. It is simple to check that

$$\min_i(a_i+b_i)=a=\left\lfloor \sqrt{4N/3}\right\rfloor,$$

verifying the condition of the lemma. □

Remark 4.3. One can verify by computer that for all cashew primes $p<150000$, the number of p-Farey triples $(a_0,b_0,a_1,b_1,a_2,b_2)$ verifying the cashew property is bounded by 2 up to permutation and reflection. We expect that this property possibly holds for all cashew primes.

Let us now establish the main result of this section.

Proof (of part 4 in Theorem 1.3). Let us first make the following computational observations; see also Table 1 and part 1 of the theorem for some of the exact values.

- If $N\in\{4,5,9,11,17,23,25\}$, then N is cashew and $m(\Gamma_0(N))=\left\lfloor \sqrt{4N/3}\right\rfloor$.
- If $N\in\{2,3,7,13,19,29,31\}$, then N is not cashew and $m(\Gamma_0(N))=\lfloor\sqrt{N}\rfloor<\left\lfloor \sqrt{4N/3}\right\rfloor$.

Hence, in the case when $N=p<37$ or $N=p^2<37$ for a prime p, we see that N is cashew if and only if and $m(\Gamma_0(N))=\lfloor\sqrt{N}\rfloor<\left\lfloor \sqrt{4N/3}\right\rfloor$. We may now consider only the case $N\geqslant 37$; furthermore, we assume $N=p$, or $N=p^2$ for a prime p, or $N=pq>81\cdot 83$ for primes p and q satisfying

$$\sqrt{p}<\sqrt{q}<\sqrt{p}+\sqrt{2}.$$

In particular, we have that

$$m(\Gamma_0(N))\leqslant\left\lfloor \sqrt{4N/3}\right\rfloor>\sqrt{N}.$$

For the backward direction of the theorem, we let s,t,a,b be as in Definition 4.1. By the observation above, we have $b>0$. There exists a maximal $\Gamma_0(N)$–polygon P such that the denominator sequence dX of the cusp set of P satisfies $\max dX=m(\Gamma_0(N))\leqslant a$. In the proof of Lemma 4.2, we have constructed a Farey N–triple $\{(a_i,b_i)\}$ containing (s,t) such that

$$s+t>a=\min_i(a_i+b_i)=\left\lfloor \sqrt{4N/3}\right\rfloor>\sqrt{N}.$$

By Lemma 3.7, this triple corresponds to free sides in Farey_v^* $\max_i\max(a_i,b_i)<\sqrt{N}$. Lemma 2.5 implies that neither s^2+t^2 nor s^2+st+t^2 are multiples of N.

Claim 1. *We have that (s,t) is a side of dX.*

Let $(\bar{s},\bar{t})$ be the Farey pair corresponding to (s,t) by Lemma 2.6. Assume for contradiction that the geodesic $\gamma := (\bar{s},\bar{t}) \subseteq \mathscr{H}$ is not a side of P. Since $s+t > m(\Gamma_0(N))$, the exterior of P contains the point $\bar{s}\oplus\bar{t}$. It follows that γ is contained in the exterior of P, possibly except for the endpoints. We can find a side $(\bar{c},\bar{d})$ of P enclosing this geodesic γ in $\mathscr{H}$ along with the x–axis such that the corresponding side (c,d) of dX satisfies

$$c \leqslant s, \qquad d \leqslant t, \qquad c+d \leqslant \max(s,t) \leqslant v.$$

Since $(c+d)^2 \leqslant N$, we see that (c,d) is neither even nor odd, and that another side (α,β) of dX is paired to (c,d) such that $\alpha,\beta \leqslant \max dX \leqslant a$. For some $k \geqslant 1$, we see from the conditions in Definition 4.1 that

$$k(sa+tb) = kN = \alpha c + \beta d \leqslant a \cdot \max(s,t) \leqslant \max(sa, t(b+s)) < sa+tb = N.$$

This is a contradiction, and so Claim 1 follows.

Claim 2. *We have that either (a,b) or $(a-t,a)$ is a side of dX.*

Suppose a side (α,β) of dX is paired to (s,t). For some $k \geqslant 1$ we again see as above that

$$N \leqslant kN = \alpha s + \beta t \leqslant 2\max(s,t)a < 2N.$$

Therefore, $k=1$ and $s\alpha + t\beta = sa+tb$. This equation has solutions of the form $(\alpha,\beta) = (a+tq, b-sq)$, where $q \in \mathbb{Z}$. In the case when $s > a-b$, the equation has a unique solution $(\alpha,\beta) = (a,b)$ satisfying $0 < \alpha,\beta \leqslant a$. This implies that (a,b) is a side in dX paired with (s,t). If $s = a-b$, then the equation has two solutions $(\alpha,\beta) = (a,b)$ and $(\alpha,\beta) = (a-t, b+s=a)$ satisfying $0 < \alpha,\beta \leqslant a$; in this case $(a-t,a)$ or (a,b) is a side paired with (s,t). Claim 2 is proved.

The above claim immediately implies the conclusion since

$$\left\lfloor \sqrt{4N/3} \right\rfloor = a \leqslant \max dX = m(\Gamma_0(N)) \leqslant \left\lfloor \sqrt{4N/3} \right\rfloor.$$

Let us now consider the forward direction. We enumerate the collection of all the N–Farey triples among the free sides in $d\,\mathrm{Farey}_v^*$ as

$$\sqcup_{1\leqslant i\leqslant k}\{(a_j^i, b_j^i)\colon\ j=0,1,2\}.$$

When $N=p$ or $N=p^2$, we see from Lemma 3.7 that

$$\left\lfloor \sqrt{4N/3} \right\rfloor = m(\Gamma_0(N)) \leqslant \max_i \min_j (a_j^i + b_j^i) \leqslant \left\lfloor \sqrt{4N/3} \right\rfloor.$$

When $N=pq$, we also have

$$\left\lfloor \sqrt{4N/3} \right\rfloor = m(\Gamma_0(N)) \leqslant \max\left(q, \max_i \min_j (a_j^i + b_j^i)\right) \leqslant \max\left(q, \left\lfloor \sqrt{4N/3} \right\rfloor\right) = \left\lfloor \sqrt{4N/3} \right\rfloor.$$

In all cases, we conclude that N is cashew by Lemma 4.2 □

Remark 4.4. In the above inequalities, we note that the collection $\{(a_j^i, b_j^i)\}$ is nonempty. If $N=p$ or $N=p^2$ with $N \geqslant 37$, then this follows from that $\left\lfloor \sqrt{N} \right\rfloor < \left\lfloor \sqrt{4N/3} \right\rfloor$. In the case when $N=pq$ with $N > 81\cdot 83$, the nonemptiness is a consequence of $q \leqslant \left\lfloor \sqrt{4N/3} \right\rfloor$.

Acknowledgements

The authors are very grateful to Asbjørn Nordentoft for his interest in this work and for pointing us to the paper [9]. The authors are also thankful to Yichen Tao for his help with some coding of initial versions of the algorithm for finding the special polygons. The first named author is supported by the Institute of Mathematics, Vietnam Academy of Science and Technology under the code IM-VAST01-2022.01. The second named author is supported by Samsung Science and Technology Foundation under Project Number SSTF-BA1301-51. The last named author is supported by the National University of Singapore academic research grant A-8000989-00-00.

References

1. Viktor Bouniakowsky, *Sur les diviseurs numériques invariables des fonctions rationnelles entières*, Mém. Acad. Sc. St. Pétersbourg. **6** (1857), 305–329.
2. Shih-Ping Chan, Mong-Lung Lang, Chong-Hai Lim, and Ser Peow Tan, *Special polygons for subgroups of the modular group and applications*, Internat. J. Math. **4** (1993), no. 1, 11–34. MR1209958
3. Yasuhiro Chuman, *Generators and relations of* $\Gamma_0(N)$, J. Math. Kyoto Univ. **13** (1973), 381–390. MR348001
4. H. Iwaniec, *Primes represented by quadratic polynomials in two variables*, Acta Arith. **24** (1973/74), 435–459. MR342464
5. Ha Huy Khoai, *Sur les séries L associées aux formes modularies*, Bull. Soc. Math. France **120** (1992), no. 1, 1–13. MR1146784
6. Ravi S. Kulkarni, *An arithmetic-geometric method in the study of the subgroups of the modular group*, Amer. J. Math. **113** (1991), no. 6, 1053–1133. MR1137534
7. Hans Rademacher, *Über die erzeugenden von kongruenzuntergruppen der modulgruppe*, Abh. Math. Sem. Univ. Hamburg **7** (1929), no. 1, 134–148. MR3069525
8. Goro Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publications of the Mathematical Society of Japan, vol. 11, Princeton University Press, Princeton, NJ, 1994, Reprint of the 1971 original, Kanô Memorial Lectures, 1. MR1291394
9. Raphael S. Steiner, *Small diameters and generators for arithmetic lattices in* $\mathrm{SL}_2(\mathbb{R})$ *and certain Ramanujan graphs*, preprint (2022).
10. Arnold Walfisz, *Weylsche Exponentialsummen in der neueren Zahlentheorie*, Mathematische Forschungsberichte, XV, VEB Deutscher Verlag der Wissenschaften, Berlin, 1963. MR0220685
11. D. Zagier, *Modular parametrizations of elliptic curves*, Canad. Math. Bull. **28** (1985), no. 3, 372–384. MR790959

Institute of Mathematics, Vietnam Academy of Science and Technology, Vietnam
Email address: dnminh@math.ac.vn

School of Mathematics, Korea Institute for Advanced Study (KIAS), Seoul, Korea
Email address: skim.math@gmail.com

Singapore
Email address: lang2to46@gmail.com

Department of Mathematics, National University of Singapore, Singapore
Email address: mattansp@nus.edu.sg