Boletín de la Sociedad Matemática Mexicana Vol. 39, 1994

FORD POLYGONS FOR $\Gamma_0(N)$

BY ANTONIO LASCURAIN ORIVE

1. Introduction

The set of matrices in the classical modular group $SL(2,\mathbb{Z})$ which are of the form $\binom{a\ \ b}{kN\ t}$, $k \in \mathbb{Z} - \{0\}$, define the Hecke congruence subgroup of level N. These finite index subgroups, denoted by $\Gamma_0(N)$, $N \in \mathbb{N}$, play an important role in number theory. The description of fundamental domains for subgroups of the modular group might turn out to be a useful geometric tool in number theoretical problems; it may also give some insight in the study of the structure of the rings $\mathbb{Z}_N$. Fundamental regions for these groups have been investigated in the simplest cases, that is, when N is a prime number (see [3] or [9]). A more general study was done recently by Kulkarni [4], who constructed fundamental domains for $\Gamma_0(N)$; these polygons have the least number of sides; however they are not Ford domains.

In this paper we study the Ford fundamental polygon for $\Gamma_0(N)$, which we will denote by R_N. First, parabolic and elliptic vertices are discussed: it is shown that their cardinality, location, and distribution in cycles are closely related to the prime factorization of N (Theorems 1 and 2, Corollaries 1 and 2, Propositions 1, 2 and 3). Another result describes the shape of these polygons: R_N looks basically like the union of ρ copies of $R_{\overline{N}}$, where $\overline{N}$ denotes the square free part of N and $\rho = N/\overline{N}$ (Theorem 3). Further information on these polygons can be found in [5].

2. Preliminaries

We denote by $\overline{\Gamma}_0(N)$ the group of transformations defined by $\Gamma_0(N)$. Accordingly, if a matrix g belongs to $\Gamma_0(N)$, $\overline{g}$ will be the corresponding transformation.

This group acts on

$$\mathbb{H}^2 = \{z \in \mathbb{C} \mid \operatorname{Im} z > 0\},$$

as hyperbolic isometries. The transformation $z \mapsto z+1$ generates the subgroup of translations; a fundamental domain for this subgroup is the set

$$R_\infty = \{z \in \mathbb{H}^2 \mid 0 < \operatorname{Re} z < 1\}.$$

A transformation $\overline{g} \in \overline{\Gamma}_0(N)$ which is not a translation acts as a euclidean isometry on exactly one circle in the complex plane. This circle is called the isometric circle of $\overline{g}$ (or g); we will denote it by $I(\overline{g})$ (or $I(g)$).

Analytically, if the matrix $g = \binom{a\ \ b}{kN\ t} \in \Gamma_0(N)$, the isometric circle of g is given by the equation

$$\{z \in \mathbb{C} \mid |g'(z)| = 1\}.$$

This is the euclidean circle $\{z \in \mathbb{C} \mid |kNz+t| = 1\}$, which has radius $1/kN$ and center $-t/kN$. Hence, the radii of isometric circles of transformations of $\overline{\Gamma}_0(N)$ are always numbers of the form $1/kN$, $k \in \mathbb{N}$. Moreover, the determinant condition implies that the centers of these circles are precisely the rational points t/kN, where $t \in \mathbb{Z} - \{0\}$, $k \in \mathbb{N}$, and t is relatively prime to kN.

A fundamental property of isometric circles states that the transformation $\overline{g}$ sends the unbounded component of $\hat{\mathbb{C}} - I(g)$, the one containing ∞, onto the bounded component of $\hat{\mathbb{C}} - I(g^{-1})$. In particular, $\overline{g}(I(g)) = I(g^{-1})$. These results follow from the chain rule; we write ext $I(g)$ and int $I(g)$ to denote such components.

The isometric circle of a transformation $\overline{g} \in \overline{\Gamma}_0(N)$ is orthogonal to the real axis, and therefore contains a geodesic in $\mathbb{H}^2$ that is also called the isometric circle of $\overline{g}$ (or g). In order to avoid a cumbersome notation, the symbols $I(\overline{g})$ and $I(g)$ will also denote the geodesic determined by the euclidean circle.

The Ford fundamental domain R_N for $\Gamma_0(N)$ defined on the infinite strip R_∞ is the hyperbolic polygon

$$R_N = R_\infty \cap \bigcap_{\overline{g}} \operatorname{ext} I(\overline{g}),$$

where $\overline{g}$ runs over all transformations in $\overline{\Gamma}_0(N)$ which are not translations.

Since $\Gamma_0(N)$ is finitely generated, R_N is a locally finite convex fundamental polygon for $\overline{\Gamma}_0(N)$, bounded by two vertical lines and a finite number of arcs. Cf. [1], chapter 9 and 10, or [2]. This polygon is also symmetric with respect to the line

$$\left\{ z \in \mathbb{H}^2 \mid \operatorname{Re} z = \frac{1}{2} \right\}.$$

This follows because the matrix $\left(\begin{smallmatrix} * & * \\ kN & -t \end{smallmatrix}\right) \in \Gamma_0(N)$ if and only if the matrix $\left(\begin{smallmatrix} * & * \\ kN & -(kN-t) \end{smallmatrix}\right) \in \Gamma_0(N)$.

A side of R_N that is contained in a circle of radius $1/kN$, $k \in \mathbb{N}$, will be called a k-side. Therefore, the sides of R_N are either k-sides or vertical lines.

The signature of $\overline{\Gamma}_0(N)$ is well known: First, if a number N has prime decomposition $2^r p_1^{r_1} \dots p_m^{r_m}$, where $r = 0, 1$ and $p_j \equiv 1 \bmod 4$ for all $j \in \{1, \dots, m\}$, then the number of conjugacy classes of elliptic subgroups of order two in $\overline{\Gamma}_0(N)$ is 2^m. However, if N has a different prime decomposition, $\overline{\Gamma}_0(N)$ has no elliptic elements of order two. In both cases the number of such classes is denoted by $v_2(N)$, or simply v_2.

Similarly, if a number N has prime decomposition $3^r p_1^{r_1} \dots p_m^{r_m}$, where $r = 0, 1$ and $p_j \equiv 1 \bmod 3$ for all $j \in \{1, \dots, m\}$, then the number of conjugacy classes of subgroups of order three in $\overline{\Gamma}_0(N)$ is 2^m. However, if N has a different prime decomposition, $\overline{\Gamma}_0(N)$ does not have elements of order three. In both cases the number of such classes is denoted by $v_3(N)$, or simply v_3. Cf. [7] or [8].

The action of $\overline{\Gamma}_0(N)$ on the Riemann sphere defines an equivalence relation on the set of parabolic fixed points. An equivalence class is called a cusp; the number of such classes is given by

$$(*) \qquad \sum_{\substack{d \mid N,\\ d>0}} \phi\left(\left(d, \frac{N}{d}\right)\right),$$

where $\left(d, \frac{N}{d}\right)$ denotes the greatest common divisor of d and N/d, and ϕ, the Euler function. $\phi(n)$ is the number of elements in $\{1, 2, 3, \ldots, n\}$ which are relatively prime to n. The number of cusps of $\overline{\Gamma}_0(N)$ $(*)$ is denoted by $v_\infty(N)$, or simply v_∞.

Finally, the genus of the Riemann surface $\mathbb{H}^2/\overline{\Gamma}_0(N)$ is given by

$$1 + \frac{\mu}{12} - \frac{v_2}{4} - \frac{v_3}{3} - \frac{v_\infty}{2},$$

where μ denotes the index of $\Gamma_0(N)$ in $SL(2, \mathbb{Z})$:

$$\mu = N \prod_{p \mid N} (1 + p^{-1}).$$

Cf. [7] or [8]. Another proof of the formula for cusps $(*)$ appears in section 3.

3. Parabolic vertices

The equivalence classes of parabolic vertices of R_N defined by the action of $\overline{\Gamma}_0(N)$ are called parabolic cycles; observe that the number of parabolic cycles is precisely v_∞. The location and cardinality of parabolic vertices of R_N is determined by the prime decomposition of N. It turns out that the number of these vertices is given in terms of the Euler function (Theorem 1). Moreover, the parabolic cycles are naturally related to the divisors of N, in the sense that the coordinates and the cardinality of a cycle depend on a fixed divisor d of N (Theorem 2 and Corollary 1). The methods of Theorem 2 also lead to a proof of the formula for cusps of $\overline{\Gamma}_0(N)(*)$, which is based only on the geometry of isometric circles and elementary number theory.

LEMMA (1). *If α is a parabolic vertex of R_N which is not the point at ∞, then α is of the form t/N, where $t \in \{0, 1, \ldots, N\}$.*

Proof. Suppose that the parabolic vertex α is not of the form stated, so $\alpha = t/kN$ where $(t, k) = 1$, $k > 1$. Take s, m such that $mt - sk = 1$. If $(s, N) = 1$, then there is a matrix $g = \left(\begin{smallmatrix} * & * \\ mN & -s \end{smallmatrix}\right) \in \Gamma_0(N)$, and since

$$\left| \frac{s}{mN} - \alpha \right| = \frac{1}{mkN} < \frac{1}{mN},$$

we have $\alpha \in \operatorname{int} I(g)$, contradicting the fact that α is a vertex. If $(s, N) > 1$, the matrix $f = \left(\begin{smallmatrix} * & * \\ m'N & -s' \end{smallmatrix}\right) \in \Gamma_0(N)$, where $m' = m + \dfrac{Nk}{(N, s)}$ and $s' = s + \dfrac{Nt}{(N, s)}$, and since $m't - s'k = 1$, we have that $\alpha \in I(f)$, which is again a contradiction. ■

Before studying parabolic vertices we make some useful remarks concerning isometric circles of transfomations in $\overline{\Gamma}_0(N)$.

1. *A rational point of the form t/N, $t \in \mathbb{Z}$, is not the center of an isometric circle of radius smaller than* $1/N$. This follows because if the isometric circle defined by a matrix $g = \left(\begin{smallmatrix} * & * \\ kN & s \end{smallmatrix}\right) \in \Gamma_0(N)$ has center t/N, then $t = -s/k$ and therefore $k|s$; but since k and s are relatively prime, we have that $k = 1$.
2. *A rational point of the form t/kN, $t \in \mathbb{Z}$, $k \in \mathbb{N}$ lies in the euclidean closure of the exterior of any circle with center s/kN, $s \in \mathbb{Z}$, $s \neq t$, and radius $1/kN$.*
3. *Rational points of the form t/N lie in the euclidean closure of the exterior of any isometric circle of radius smaller than* $1/N$. This is a consequence of remarks 1 and 2.
4. *The isometric circles of radius smaller than $1/N$ are contained in the region $\{z \in \mathbb{C} | \operatorname{Im} z \leq 1/2N\}$.*
5. *An isometric circle of radius $1/N$ which intersects R_∞ is visible; that is, it contains a side of R_N. In particular, since the matrix $g = \left(\begin{smallmatrix} * & * \\ N & -1 \end{smallmatrix}\right)$ belongs to $\Gamma_0(N)$, 1-sides always exist.* This is clear from remark 4.

THEOREM (1). *R_N has $N - \phi(N) + 2$ parabolic vertices, where ϕ denotes the Euler function. These vertices are rational points of the form t/N, where $t \in \{0, 1, \ldots, N\}$, and $(t, N) > 1$, together with the point at infinity.*

Proof. It follows from Lemma 1 that a parabolic vertex is a point of the form t/N, $t \in \{0, 1, \ldots, N\}$, or the point at infinity. We claim that among these rational points only those for which $(t, N) > 1$ are parabolic vertices.

First, if $(t, N) = 1$, the matrix $\left(\begin{smallmatrix} * & * \\ N & -t \end{smallmatrix}\right) \in \Gamma_0(N)$, and therefore t/N is not visible.

However, if $(t, N) > 1$, we may find $k \in \mathbb{N}$ such that the matrix

$$\begin{pmatrix} * & * \\ kN & -(kt + 1) \end{pmatrix} \in \Gamma_0(N),$$

for instance, $k = N/(t, N)$. Among these matrices let g denote the one defined by the smallest k, that is, the one with largest isometric circle. Observe that $t/N \in I(g)$. Similarly we may take the smallest integer m for which the matrix

$$f = \begin{pmatrix} * & * \\ mN & -(mt - 1) \end{pmatrix} \in \Gamma_0(N).$$

We also have $t/N \in I(f)$.

Now, since $(t, N) > 1$, the remarks 1, 2 and 3 imply that t/N lies in the euclidean closure of the exterior of any isometric circle of a transformation in $\overline{\Gamma}_0(N)$. Hence, $I(g)$ and $I(f)$ contain visible sides of R_N that end at the point t/N, which is therefore a parabolic vertex. This follows because the other isometric circles which contain the point t/N are tangent to either $I(g)$ or $I(f)$ in their interior.

Consequently, there are $N - 1 - \phi(N)$ parabolic vertices of the form t/N, $t \in \{1, 2, \ldots, N-1\}$, $(t, N) > 1$. The remarks 3 and 5 also show that 0 and 1 are parabolic vertices. Finally, Lemma 1 implies that there are no other vertices apart from the point at infinity. ■

The parabolic cycles of R_N are determined by the divisors of N in the following way.

THEOREM (2). *Two parabolic vertices t_1/N, t_2/N of R_N are $\overline{\Gamma}_0(N)$ equivalent if and only if*

$$(t_1, N) = (t_2, N),$$

and

$$t_1/d \equiv t_2/d \mod (d, N/d),$$

where $d = (t_i, N), i = 1, 2$.

Proof. Suppose first that t_1/N and t_2/N are two $\overline{\Gamma}_0(N)$ equivalent parabolic vertices, $t_1 < t_2$, and let $g \in \Gamma_0(N)$ be such that $\overline{g}(t_1/N) = t_2/N$. The properties of isometric circles imply that $t_1/N \in I(g)$. This follows because if $t_1/N \in \text{Ext}I(g)$. then $t_2/N \in \text{Int}(g^{-1})$ and t_2/N would not be visible; for the same reason t_1/N is not in the interior of $I(g)$.

Hence, since $\overline{g}$ preserves orientation and $\overline{g}(I(g)) = I(g^{-1})$, the matrix g can be expressed as

$$\begin{pmatrix} kt_2 + 1 & * \\ kN & -(kt_1 - 1) \end{pmatrix}$$

or

$$\begin{pmatrix} kt_2 - 1 & * \\ kN & -(kt_1 + 1) \end{pmatrix}.$$

This situation is described in Figure 1 for g hyperbolic.

In the first case, since $\det g = 1$, we have that

$$-k^2 t_2 t_1 + k(t_2 - t_1) \equiv 0 \mod kN.$$

this is equivalent to

$$(1) \qquad t_2 - t_1 \equiv kt_1t_2 \mod N,$$

and so $(t_1, N) = (t_2, N)$.

To prove the second part we may write (1) as follows,

$$\frac{t_2}{d}-\frac{t_1}{d}=n\frac{N}{d}+kd\left(\frac{t_1}{d}\right)\left(\frac{t_2}{d}\right),$$

where $d=(t_i,N)$, $i=1$, 2, and n is an integer.
Thus

$$\frac{t_2}{d}\equiv\frac{t_1}{d}\bmod\left(\frac{N}{d},d\right).$$

A similar argument holds in the second case.

Viceversa, given t_1, $t_2\in\mathbb{Z}$, $t_1<t_2$, such that $(t_1,N)=(t_2,N)$ and $\frac{t_2}{d}\equiv\frac{t_1}{d}\bmod w$, where $d=(t_i,N)$, $i=1$, 2, and $\left(\frac{N}{d},d\right)=w$, we prove that t_1/N, t_2/N are $\overline{\Gamma}_0(N)$ equivalent.

Since $\left(\frac{t_i}{d},\frac{N}{d}\right)=1$, $i=1,2$, we have that

$$\left(\frac{t_1}{d}\frac{t_2}{d}d,\frac{N}{d}\right)=\left(\frac{N}{d},d\right).$$

Therefore we may write

$$w=qd\left(\frac{t_1}{d}\right)\left(\frac{t_2}{d}\right)+m\frac{N}{d},\quad q,m\in Z, \tag{2}$$

and reverse the steps above by putting $\frac{t_2}{d}-\frac{t_1}{d}=wu$, $u\in\mathbb{N}$, and inserting (2) in this last expression.

If $q>0$, putting $k=uq$ we get a matrix

$$\begin{pmatrix}kt_2+1 & um\\ kN & -(kt_1-1)\end{pmatrix}\in\Gamma_0(N),$$

whose corresponding transformation sends t_1/N to t_2/N.

If $q<0$ and $k=-uq$, we get

$$\begin{pmatrix}kt_2-1 & -um\\ kN & -(kt_1+1)\end{pmatrix}\in\Gamma_0(N).$$

■

We remark that the sufficiency part of Theorem 2 does not assume that t_1/N and t_2/N are parabolic vertices, hence the statement is more general. In particular, any two rational points t_1/N and t_2/N such that $(t_i,N)=1$, $i=1$, 2 are $\overline{\Gamma}_0(N)$ equivalent.

Theorem 2 also leads us to a new proof of the formula (∗) for the number of cusps of $\overline{\Gamma}_0(N)$, which is somewhat more geometric that those in [7] and [8].

The proof goes like this: To each proper divisor d of N we associate all parabolic vertices in R_N of the form t/N, where $(t, N) = d, t \in \{2, 3, \ldots, N-2\}$. Therefore, it follows from Theorem 2 that if two parabolic vertices define the same cusp, then they are associated to the same divisor. Observe that the set of all parabolic vertices associated with a divisor d is in one to one correspondence with the subset of numbers in $\{1, 2, \ldots, N/d\}$ which are coprime with N/d, and so there are $\phi(N/d)$ of these vertices. It is clear that any of these vertices t/N associated to a divisor d is uniquely determined by a number $m \in \{1, 2, \ldots, N/d\}$, where $t = dm$.

Now, it is convenient to decompose N/d as nq, where q denotes the biggest factor of N/d which is coprime with $\left(\frac{N}{d}, d\right)$; so $n = N/qd$, and all the primes which are divisors of n are also divisors of $\left(\frac{N}{d}, d\right)$. With this notation the numbers $\{1, 2, \ldots, N/d\}$ may be enumerated as follows:

$$\begin{array}{llll}
1, & 2, & \ldots, w, & w+1, \ldots, \; n \\
n+1, & n+2, & \ldots, n+w, & \qquad \ldots, 2n \\
\vdots & \vdots & \vdots & \qquad \vdots \\
(q-1)n+1, & (q-1)n+2, & \ldots, (q-1)n+w, & \qquad \ldots, qn,
\end{array}$$

where $w = \left(\frac{N}{d}, d\right)$.

Since $\phi(N/d) = \phi(n)\phi(q)$, there are exactly $\phi(n)$ columns, each containing $\phi(q)$ numbers that represent parabolic vertices associated with the divisor d. Moreover, since any two numbers in the same column are congruent mod n and therefore mod w, Theorem 2 implies that the parabolic vertices associated with the numbers in a fixed column define the same cusp. This Theorem also shows that among the first w columns, $\phi(w)$ of them represent parabolic vertices in $\phi(w)$ different cusps; furthermore, it says that the other columns do not define new cusps. Hence, the formula for cusps $(*)$ follows. The cusps defined by the cycles $\{\infty\}$ and $\{0, 1\}$ are, of course, those associated to the divisors 1 and N.

We may also count the number of parabolic vertices in each cycle; since $\phi(n) = \frac{n}{w}\phi(w)$, each cycle has $\frac{n}{w}\phi(q)$ parabolic vertices. This leads to our next result.

COROLLARY (1) *The length of a parabolic cycle in R_N defined by a divisor d as described above, is $\frac{\phi(N/d)}{\phi(w)}$, where $w = \left(d, \frac{N}{d}\right)$.*

As an example we show the parabolic vertices for $N = 18$ (see figure 2):

For $d = 2$, we have $N/d = 9$, $w = 1$, one cusp with parabolic vertices at $t/18$, $t = 2, 4, 8, 10, 14, 16$.

For $d = 3$, we have $N/d = 6$, $w = 3$, two cusps with vertices at $3/18$ and $15/18$.

For $d = 6$, we have $N/d = 3$, $w = 3$, two cusps with vertices at $6/18$ and $12/18$.

For $d = 9$, we have $N/d = 2$, $w = 1$, one cusp with vertex at $9/18$.

We finish this section by remarking that the proof of the formula for the number of cusps $(*)$ given in [7] is not complete. The last part of the argument is false; for instance, it fails for the particular case $N = 35$ and $d = 5$.

4. Elliptic vertices

Since $\overline{\Gamma}_0(N)$ is a subgroup of the modular group of transformations $PSL(2, \mathbb{Z})$, all elliptic elements are of order two or three. As we mentioned in the preliminaries, the number of elliptic classes of order two is well known; it is a power of 2 if -1 is a quadratic residue of N and it is zero otherwise. Similarly there are classes of order three if and only if -3 is a quadratic residue of N (see [7] or [8]).

In this section we count the number of elliptic vertices of R_N and describe their locations. We first show the coordinates of elliptic vertices of order two; it turns out that their imaginary part is $1/N$, and therefore they are above all the isometric circles. In particular, the corresponding cycles have length one. The cardinality is obtained by establishing a bijection between these vertices and the solutions in $\mathbb{Z}_N$ to the equation $t^2 = -1$.

PROPOSITION (1). *The elliptic vertices of order two in R_N are precisely the points $\dfrac{t}{N} + \dfrac{i}{N}$, where $1 < t < N$, and $t^2 \equiv -1 \bmod N$. In particular, each corresponding cycle has length one.*

Proof. We may assume that $N = 2^r p_1^{r_1} \dots p_m^{r_m}$, where $r = 0, 1$, and $p_j \equiv 1 \bmod 4$, for all $j \in \{1, \dots, m\}$. Now, given $t \in \{1, 2, \dots, N - 1\}$ such that $t^2 \equiv -1 \bmod N$, one gets a matrix

$$\begin{pmatrix} t & * \\ N & -t \end{pmatrix} \in \Gamma_0(N),$$

whose corresponding transformation is elliptic of order two, with fixed point $\dfrac{t}{N} + \dfrac{i}{N}$. These points are visible since their imaginary parts are greater than or equal to all other points contained in the isometric circles of $\overline{\Gamma}_0(N)$. Furthermore, as the angle sum of a cycle of elliptic vertices of order two is π (see [1], Theorem 9.3.5), these points belong to cycles of length one. This is a consequence of remark 4 in section 3, because a vertex of order two is contained in exactly one isometric circle of $\Gamma_0(N)$.

The result now follows since Euler's criterion and the theory of quadratic residues show that the number of solutions in $\mathbb{Z}_N$ to the equation $t^2 = -1$ is 2^m. Cf. [6], Theorems 5.1 and 5.2. ∎

A simple calculation shows that if a pair of matrices of the form $\left(\begin{smallmatrix} t & * \\ k_1N & -t\end{smallmatrix}\right)$, $\left(\begin{smallmatrix} s & * \\ k_2N & -s\end{smallmatrix}\right)$, $k_1, k_2 \in \mathbb{Z} - \{0\}$ are conjugate in $\Gamma_0(N)$, then $t \equiv s \bmod N$. This fact in conjunction with Proposition 1 lead us to describe the location and distribution in equivalence classes of the elliptic fixed points of order two defined on $\mathbb{H}^2$ by the action of $\overline{\Gamma}_0(N)$. We recall that given a transformation $g \in PSL(2, \mathbb{R})$ of order two, the fixed point of $\overline{g}$ is the highest point of the isometric circle $I(\overline{g})$; that is, it has greater imaginary part than other points in $I(\overline{g})$. Thus, for the case of $\Gamma_0(N)$ these points have coordinates of the form

$$\frac{s}{kN} + \frac{i}{kN}, \quad s \in \mathbb{Z}, k \in \mathbb{N}.$$

PROPOSITION (2). *Let $\alpha \in \mathbb{H}^2$ be the elliptic fixed point of a transformation $\overline{g} \in \overline{\Gamma}_0(N)$ of order two. Then α is $\overline{\Gamma}_0(N)$ equivalent to the vertex $\frac{t}{N} + \frac{i}{N} \in R_N$ if and only if $t \equiv s \bmod N$, where $\alpha = \frac{s}{kN} + \frac{i}{kN}$.*

Proof. Let $\overline{f}$ denote the transformation fixing $\frac{t}{N} + \frac{i}{N}$. Then $\overline{f}$ and $\overline{g}$ are defined by matrices in $\Gamma_0(N)$ of the form $f = \left(\begin{smallmatrix} t & b \\ N & -t\end{smallmatrix}\right)$ and $g = \left(\begin{smallmatrix} s & b' \\ kN & -s\end{smallmatrix}\right)$, $k \in \mathbb{N}$. Now, if the matrix $h = \left(\begin{smallmatrix} a & \beta \\ nN & -u\end{smallmatrix}\right) \in \Gamma_0(N)$ defines the transformation $\overline{h}$ for which $\overline{h}\left(\frac{t}{N} + \frac{i}{N}\right) = \frac{s}{kN} + \frac{i}{kN}$, then $\overline{g} = \overline{h}\overline{f}\overline{h}^{-1}$. In terms of matrices this means that either $g = hfh^{-1}$ or $-g = hfh^{-1}$; however, we claim that the second possibility does not happen. To prove the claim we use the relation $-ua - \beta nN = 1$ to calculate

$$hfh^{-1} = \begin{pmatrix} * & * \\ N(-2ntu + u^2 - n^2bN) & N(\text{integer}) - t \end{pmatrix}.$$

This expression shows that

$$\operatorname{Im} \alpha = \pm \frac{1}{N(2ntu + u^2 - n^2bN)}.$$

On the other hand,

$$\operatorname{Im} \alpha = \operatorname{Im} \left\{\overline{h}\left(\frac{t}{N} + \frac{i}{N}\right)\right\} = \frac{1}{N|nN(\frac{t}{N} + \frac{i}{N}) - u|^2}$$
$$= \frac{1}{N[(nt - u)^2 + n^2]}.$$

Finally, another calculation shows that

$$-2ntu + u^2 - n^2bN = (nt - u)^2 + n^2;$$

hence the claim and therefore the Proposition follow. ■

The calculation in the proof of Proposition 2 shows that if the radius of the isometric circle of a transformation of order two is given by $1/kN$, then k is the square of an integer, or the sum of two squares. Thus, elliptic fixed points of order two appear only at certain "heights".

COROLLARY (2). *The imaginary parts of the elliptic fixed points of order two defined on $\mathbb{H}^2$ by the action of $\overline{\Gamma}_0(N)$ are always numbers of the form $1/kN$, where k is the square of a number or the sum of two squares.*

Hence the distribution of these fixed points is closely related to the numerical structure of N; the same happens for elliptic fixed points of order three.

We first prove that the elliptic cycles of order three in R_N have also length one, this is done by establishing a one to one correspondence between the vertices and the square roots of -3 modulo N. First we need an easy fact:

LEMMA (2) *Given N an odd number, then one has that in $\mathbb{Z}_N$ the square roots of -3 are in one to one correspondence with the solutions of the equation $t(t+1) = -1$.*

Proof. As $N | t^2 + t + 1$ if and only if

$$N | 4t^2 + 4t + 4 = (2t+1)^2 + 3,$$

the association $t \mapsto 2t + 1$, $t \in \{1, 2, \dots, N-1\}$ is the required bijection. To check surjectivity, if $u^2 \equiv -3 \bmod N$ and u is even, write $u + N = 2t + 1$ to solve for t. ■

PROPOSITION (3). *The elliptic vertices of order three in R_N are the points $\dfrac{2t+1}{2N} + \dfrac{\sqrt{3}}{2N}i$, where $t(t+1) \equiv -1$ mod N, $1 < t < N$. Consequently, the corresponding cycles have length one.*

Proof. We may assume that $N = 3^r p_1^{r_1} \dots p_m^{r_m}$, $r = 0, 1$ and $p_j \equiv 1 \bmod 3$ for all $j \in \{1, \dots, m\}$, otherwise there are no elliptic fixed points of order three. As we mentioned in the preliminaries, for N with such a prime decomposition, the number of elliptic classes of order three of $\overline{\Gamma}_0(N)$ is 2^m.

Now, given t such that $t(t+1) \equiv -1 \bmod N$, $0 < t < N$, the matrix $g = \begin{pmatrix} t+1 & * \\ N & -t \end{pmatrix}$ defines a transformation of order three in $\overline{\Gamma}_0(N)$ fixing $\dfrac{2t+1}{2N} + \dfrac{\sqrt{3}}{2N}i$. In fact, this point is the intersection of the isometric circles of g and g^{-1}. Moreover, it is visible because the other isometric circles of radius $1/N$ are contained in the set

$$\left\{ z \in \mathbb{C} \mid \operatorname{Re} z \leq \frac{t}{N} \quad \text{or } \operatorname{Re} z \geq \frac{t+1}{N} \right\},$$

and the isometric circles of smaller radius are below the line $\operatorname{Im} z = \dfrac{1}{2N}$ (see remarks 2 and 4 in section 3).

Furthermore, as the external angle at such a point is $2\pi/3$, the corresponding cycle has length one. The Proposition now follows from Lemma 2 and the fact that the number of square roots of -3 is 2^m. Cf. [6], Theorems 3.21, 4.13, 5.1, 5.3 and 5.7. ■

Similar calculations to those of Proposition 2 establish the distribution in equivalence classes of the fixed points of order three. It turns out that two matrices $\left(\begin{smallmatrix} t+1 & * \\ k_1N & -t \end{smallmatrix}\right)$ and $\left(\begin{smallmatrix} s+1 & * \\ k_2N & -s \end{smallmatrix}\right)$, $k_1, k_2, \in \mathbb{N}$ are conjugate in $\Gamma_0(N)$, if and only if $t \equiv s \bmod N$. In particular, two elliptic fixed points of order three,

$$\frac{1}{2}\left(\frac{2t+1}{k_1N}\right) + \frac{\sqrt{3}}{2}\left(\frac{i}{k_1N}\right) \quad \text{and} \quad \frac{1}{2}\left(\frac{2s+1}{k_2N}\right) + \frac{\sqrt{3}}{2}\left(\frac{i}{k_2N}\right),$$

where k_1, $k_2 \in \mathbb{N}$ are $\overline{\Gamma}_0(N)$ equivalent if and only if

$$s \equiv t \bmod N.$$

We end this section by remarking that all the vertices of R_N which are not of order two are contained in the region

$$\left\{ z \in \mathbb{C} \mid \operatorname{Im} z \leq \frac{\sqrt{3}}{2N} \right\},$$

and therefore the accidental vertices have smaller or equal imaginary parts than those of order three.

In Figure 3 we show R_{13}, where elliptic vertices of both orders appear, namely, $\frac{5}{13} + \frac{i}{13}, \frac{8}{3} + \frac{i}{13}$ of order two, and $\frac{1}{2}\left(\frac{7}{13}\right) + \frac{\sqrt{3}}{2}\left(\frac{i}{13}\right), \frac{1}{2}\left(\frac{19}{13}\right) + \frac{\sqrt{3}}{2}\left(\frac{i}{13}\right)$ of order three.

5. Reduction to square free numbers

Here the elliptic fixed points of order two in ∂R_N will not be considered as vertices, hence the two sides ending at such points will be thought of as just one side which is paired with itself. We introduce some notation: if A is a subset of the complex plane, $\overline{A}$ will denote the euclidean closure in $\mathbb{C}$; as in the introduction, the square free part of a given number N will be denoted by $\overline{N}$ and the quotient $N/\overline{N}$ by ρ.

We will show that $\overline{R_N}$ can be obtained as the union of ρ translations of the polygon obtained by contracting $\overline{R_{\overline{N}}}$ by a factor of $1/\rho$. Our next Theorem describes this fact in a precise way.

THEOREM (3). *Let $\overline{T}^m$ denote the translation by m, $m \in \mathbb{N}$, and $\overline{M}_{1/\rho}$ contraction by $1/\rho$, then*

$$\overline{R}_N = \bigcup_{m=0}^{\rho-1} \overline{M}_{1/\rho}\overline{T}^m(\overline{R}_{\overline{N}}).$$

Proof. We show that the visible isometric circles of $\Gamma_0(N)$ can be obtained from those of $\Gamma_0(\overline{N})$ by translation followed by contraction.

To do so, let $\overline{g}$ be a pairing of $R_{\overline{N}}$ which is not a translation; $\overline{g}$ is defined by a matrix $g = \left(\begin{smallmatrix} a & b \\ k\overline{N} & t \end{smallmatrix}\right)$. Hence $\overline{g}$ pairs the sides of $R_{\overline{N}}$ contained in the circles of radius $1/kN$ centered at $-t/k\overline{N}$ and $a/k\overline{N}$ (if $a = -t$, $\overline{g}$ pairs such a side with itself).

We claim that by translating these two (or one) circles, $I(g)$ and $I(g^{-1})$, n units to the right, $n = 0, 1, \dots, \rho - 1$, and then contracting them by a factor of $1/\rho$, one gets 2ρ circles (or ρ), which are isometric circles of transformations in $\Gamma_0(N)$, and their corresponding inverses.

To prove the claim let T and M_ρ denote the matrices $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, $\left(\begin{smallmatrix} \sqrt{\rho} & 0 \\ 0 & \sqrt{1/\rho} \end{smallmatrix}\right)$ respectively. Observe that a transformation sending one of the transformed circles into another can be defined by a matrix of the form

$$g_n^m = M_\rho^{-1} T^m g T^{-n} M_\rho,$$

for suitable $n, m \in \{0, 1, \dots, \rho - 1\}$. These matrices may not belong to $\Gamma_0(N)$, however calculation yields

$$g_n^m = \begin{pmatrix} a + mk\overline{N} & [-an + b - m(nk\overline{N} - t)]\rho^{-1} \\ k\rho\overline{N} & -k\overline{N}n + t \end{pmatrix};$$

hence they belong to such a group provided the upper right entry is an integer. We deduce that given an integer $n \in \{0, 1, \dots, \rho - 1\}$, there is exactly one $m \in \{0, 1, \dots, \rho - 1\}$, for which $g_n^m \in \Gamma_0(N)$. This follows because for a fixed n, the set

$$\{-an + b - m(nk\overline{N} - t)\},$$

$m \in \{0, 1, \dots, \rho - 1\}$ forms a complete set of representatives for $\mathbb{Z}_\rho$, since t and $\overline{N}$ are relatively prime.

The claim now follows because the isometric circles of the matrices g_n^m, and their inverses, are precisely the circles $M_\rho^{-1}T^n(I(g))$, $M_\rho^{-1}T^n(I(g^{-1})$, $n \in \{0, 1, \dots, \rho - 1\}$.

We remark that the claim is also true even if $I(g)$ and $I(g^{-1})$ do not contain sides of $R_{\overline{N}}$; however these geodesics must be subsets of $\overline{R}_\infty$.

Viceversa, if a matrix $f = \left(\begin{smallmatrix} a & b \\ kN & t \end{smallmatrix}\right) \in \Gamma_0(N)$ defines a pairing of sides of R_N, $\overline{f}$ pairs the isometric circles $I(f)$ and $I(f^{-1})$ of radius $1/kN$ and centers at $-t/kN$ and a/kN. We divide the unit interval [0,1] into ρ subintervals of length $1/\rho$, so there are unique numbers n and m, $0 \le n,\, m < \rho$ such that

$$\frac{n}{\rho} < \frac{-t}{kN} < \frac{n+1}{\rho}, \tag{1}$$

and

$$\frac{m}{\rho} < \frac{a}{kN} < \frac{m+1}{\rho}. \tag{2}$$

This happens since the remarks 1, 2, 3 and 5 in section 3 imply that an isometric circle of $\Gamma_0(N)$ is either contained in R_∞ or in $\overline{\mathbb{C} - R_\infty}$.

Now, the location of the centers of $I(f)$ and $I(f^{-1})$ suggests that a transformation in $\Gamma_0(\overline{N})$ derived from f might be defined by the matrix

$$h_f = T^{-m} M_\rho f M_\rho^{-1} T^n.$$

Indeed calculation shows that

$$h_f = \begin{pmatrix} a - mk\overline{N} & \text{integer} \\ k\overline{N} & t + nk\overline{N} \end{pmatrix} \in \Gamma_0(\overline{N}),$$

and (1) and (2) imply that the isometric circles of h_f and h_f^{-1} are contained in the strip R_∞. Furthermore, these circles may be obtained by expanding $I(f)$ and $I(f^{-1})$ by a factor of ρ and then translating them back to R_∞ by $-n$ and by $-m$ repectively.

Hence this is the reverse process to the claim, and thus there must be another $\rho-1$ transformations in $\Gamma_0(N)$ together with their inverses, associated to the transformations $\overline{h_f}$ and $\overline{h_f^{-1}}$ in $\Gamma_0(\overline{N})$.

Finally, an isometric circle $I(g)$ of $\Gamma_0(\overline{N})$ is visible if and only if the corresponding isometric circles $I(g_n^m)$, $n = 0, 1, \ldots, \rho - 1$ of $\Gamma_0(N)$ are visible. This follows because a family of isometric circles of $\Gamma_0(\overline{N})$ covering $I(g)$ will induce a family of isometric circles of $\Gamma_0(N)$ covering $I(g_n^m)$, $n = 0, 1, \ldots, \rho - 1$, and viceversa. ■

As an example of Theorem 3 we illustrate the cases $\overline{N} = 6$ (Figure 4) and $N = 18$ (Figure 2).

The following facts are direct consequences of this Theorem:

i) If $R_{\overline{N}}$ has s sides, then R_N has $\rho(s-2)+2$ sides.

ii) If $R_{\overline{N}}$ has s k-sides, then R_N has ρs k-sides.

iii) If $R_{\overline{N}}$ has s finite vertices, then R_N has ρs finite vertices.

iv) If $R_{\overline{N}}$ has s parabolic vertices, then R_N has $\rho(s-2)+2$ parabolic vertices.

The same repetition happens for a specific type of vertex; namely, for each vertex in $R_{\overline{N}}$ which is the intersection of m isometric circles of radii

$$1/k_1\overline{N},\ 1/k_2\overline{N},\ \ldots,\ 1/k_m\overline{N},$$

there are ρ vertices of such type in R_N.

Theorem 3 also implies that if $\overline{\Gamma}_0(\overline{N})$ is torsion free, the open Riemann surface $\mathbb{H}^2/\overline{\Gamma}_0(N)$ is a ρ-sheeted covering space of the corresponding surface $\mathbb{H}^2/\overline{\Gamma}_0(\overline{N})$. However, this is not true for the compactified surfaces; for instance, the surface defined by $N = 22$ has four cusps whereas the one defined by $N = 44$ has six cusps; this must occur because of the special role of the point at ∞. In the torsion cases the situation is more complicated; for example, $\Gamma_0(10)$ has elliptic elements whereas $\Gamma_0(20)$ has not.

I would like to thank Troels Jørgensen for suggesting to me the methods in the proof of Theorem 3. A different proof appears in [5].

All the results in this paper have their counterparts in the group

$$\Gamma^0(N) = \left\{ \begin{pmatrix} a & b \\ c & t \end{pmatrix} \in SL(2,\mathbb{Z}) \middle| b \equiv 0 \mod N \right\}.$$

Cf. [5]. This group and $\Gamma_0(N)$ are, of course, conjugate in $SL(2,\mathbb{Z})$.

Antonio Lascurain Orive
Departamento de Matemáticas
Facultad de Ciencias, UNAM
MEXICO

References

[1] A.F. Beardon, The Geometry of Discrete Groups, Springer Verlag, 1983.

[2] A.F. Beardon, and T. Jørgensen, *Fundamental Domains for Finitely Generated Kleinian Groups,* Mathematica Scandinavica, **36**, 1975.

[3] R. Fricke, Die Elliptischen Funktionen und ihre Anwendungen, part II, ch. 3, p. 349, Teubner, 1922.

[4] R. Kulkarni, *An Arithmetic-Geometric Method in the Study of the Subgroups of the Modular Group,* American Journal of Mathematics, **113**, 1991.

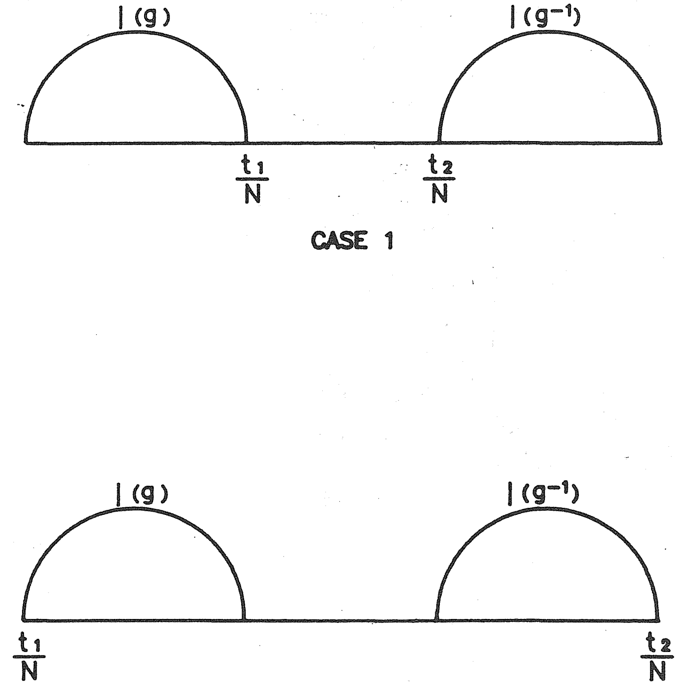
[5] A. Lascurain, Fundamental Domains for the Hecke Congruence Subgroups, Columbia University, Ph. D. thesis, 1989.

[6] W.J. LeVeque, Fundamentals of Number Theory, Addison-Wesley, 1977.

[7] G. Shimura, Introduction to the Arithmetic Theory of Automorphic Functions, Tokyo Iwanami Shoten and Princeton University Press, 1971.

[8] B. Shoeneberg, Elliptic Modular Functions, Springer Verlag, 1974.

[9] D. Zagier, *Modular Parametrizations of Elliptic Curves, Canadian Mathematical Bulletin,* **28**, 1985.



CASE 1

CASE 2

FIGURE 1

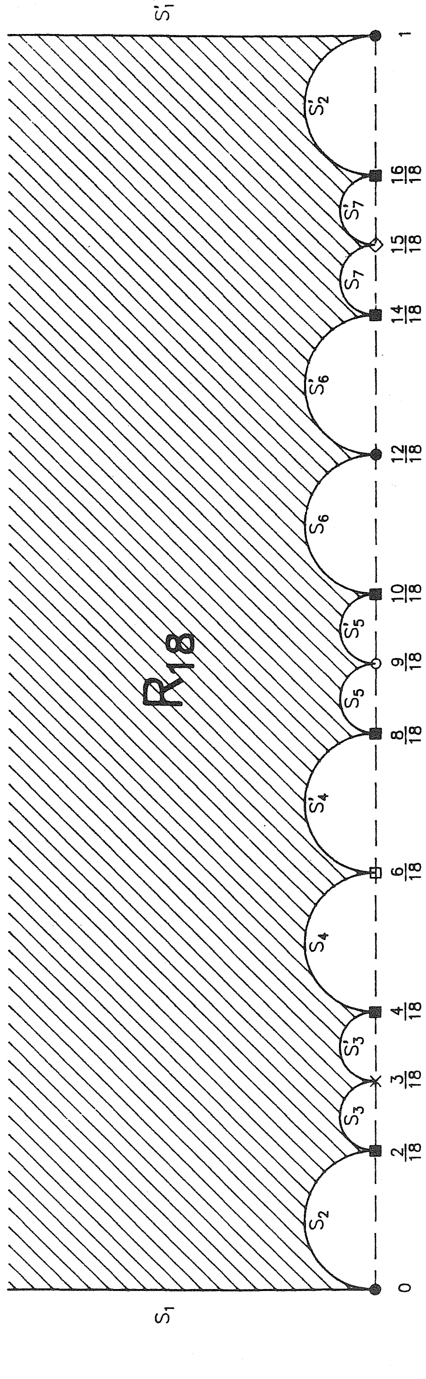

S'_1
S_1
R_18
0
S_2
2/18
S_3
3/18
S'_3
4/18
S_4
6/18
S'_4
8/18
S_5
9/18
S'_5
10/18
S_6
12/18
S'_6
14/18
S_7
15/18
S'_7
16/18
S'_2
1

FIGURE 2

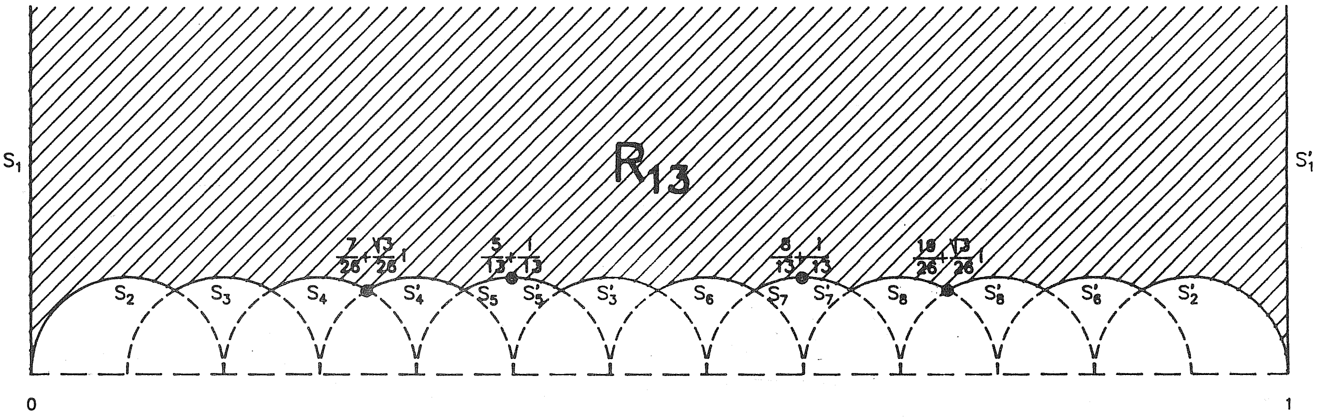

R_{13}
S_1
S_1'
$\frac{7}{26}+\frac{\sqrt{13}}{26}i$
$\frac{5}{13}+\frac{1}{13}$
$\frac{8}{13}+\frac{1}{13}$
$\frac{19}{26}+\frac{\sqrt{13}}{26}i$
S_2
S_3
S_4
S_4'
S_5
S_5'
S_3'
S_6
S_7
S_7'
S_8
S_8'
S_6'
S_2'
0
1

FIGURE 3

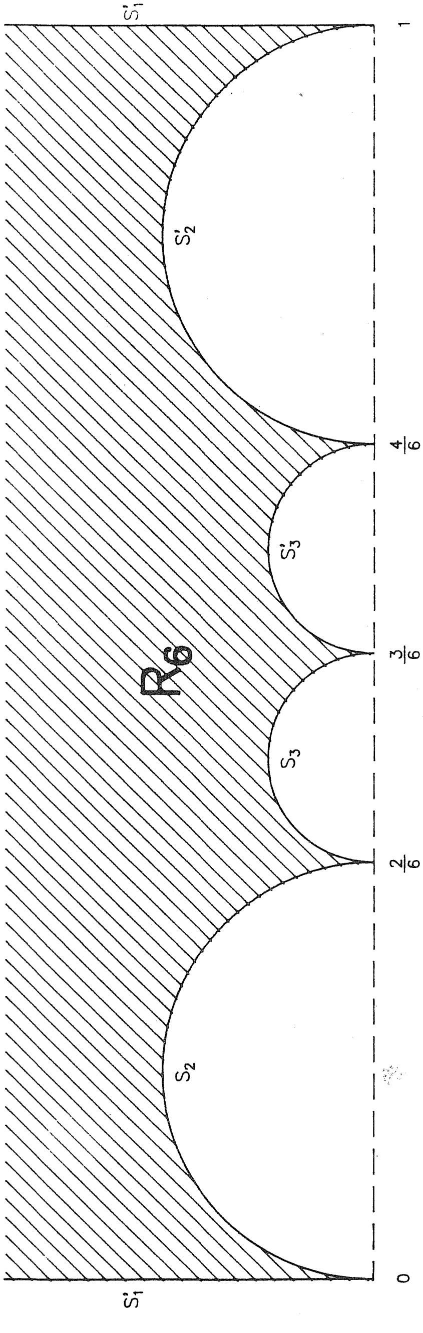


FIGURE 4