ACTA ARITHMETICA
LXIII.3 (1993)

The $\sqrt{p}$ Riemann surface

by

Mark Sheingorn (New York, N.Y.)

To My First Teacher, Jay Stepelman
On His Retirement from George Washington High School (NYC)

1. Introduction

1.1. Background. This paper is a broadening and deepening of the project begun in [7]. That paper showed that for each prime $p \equiv 1 \bmod 4$ establishing a conjecture of N. C. Ankeny, E. Artin and S. Chowla (denoted A-A-C below) about fundamental units in $\mathbb{Q}(\sqrt{p})$ was equivalent to determining which of two shapes occurred at the center of $\Gamma_p^* \backslash \mathcal{H}$. Here $\Gamma_p^* = \Gamma(p) \cap \Gamma^0(p^2)$. [7] also offered a characterization of the reflections, their pointwise fixed reflection lines, and other geometrical attributes of $\Gamma_p^* \backslash \mathcal{H}$.

The conjecture A-A-C is as follows: Let (x_0, y_0) be the fundamental solution of $x^2 - py^2 = -1$. Then $p \nmid y_0$ (see [2]). This conjecture has received a fair amount of attention for algebraic and computational number theorists. (We refer the reader to [7] for bibliographic discussion and citations.)

Briefly, what we did was effectively characterize the isometries of $\Gamma_p^* \backslash \mathcal{H}$, of which there are many, and then pay close attention to the reflections. These are of two kinds according as their pointwise fixed reflection lines—geodesics we call *reflectors*—are closed geodesics or ones emanating and terminating from cusps. The cusped ones are easier to handle, as they are essentially rational in nature; the closed ones are quadratic in nature and natural geometric questions lead to non-trivial number theoretic questions in various real quadratic extensions of $\mathbb{Q}$.

One particular closed reflection played a crucial role in the earlier work: $L_0 := z \to p/\overline{z}$. ($L_0$ is associated with the involution so central to the

The author gratefully acknowledges IMPAN and the (American) National Academy of Science for supporting a visit to Warsaw where this paper was conceived. Special thanks also to Professor Andrzej Schinzel and Dr. Zofia Adamowicz for being truly wonderful hosts.

work of Atkin and Lehner on the Hecke theory for $\Gamma^0(N)$.) The h-line $\sigma_L = (\pm\sqrt{p})$ is pointwise fixed by L_0, and thus its projection to $\Gamma_p^*\backslash\mathcal{H}$ is closed and pointwise fixed by L_0 on that surface. (We trust no confusion will arise from the fact that we will often use the same notation for reflections on $\Gamma_p^*\backslash\mathcal{H}$ and their lifts to reflections on $\mathcal{H}$.)

The connection with A-A-C is then this: *A-A-C is true or false according as σ_L is orthogonal to pointwise fixed reflection lines of one or p cusped reflections of $\Gamma_p^*\backslash\mathcal{H}$.* (We defined these reflectors as *reversed* by one another, for the obvious reason.)

(For subsequent work on the geometry of reflections of $\Gamma(N)\backslash\mathcal{H}$ and $\Gamma^0(N)\backslash\mathcal{H}$, see [1].)

1.2. The present paper. The general theme of this paper is first, that the above correspondence between the geometry of pointwise fixed reflectors and features of $\mathrm{CF}(\sqrt{p})$ continues to hold as we examine in turn: (a) more general closed reflectors, and (b) shortest connections (of differential geometry) between such reflectors. In brief, the entire $\mathrm{CF}(\sqrt{p})$ is in fact equivalent to this more general class of closed reflectors; and the incidence pattern of the connections is determined by intermediate convergents. We now turn to a statement of our results.

First, the association of L_0 with the Pell equation $x^2 - py^2 = -1$ turns out to be much more general. In fact, there are closed reflectors for each $k \bmod p$ admitting a solution of $x^2 - py^2 = k$. Here is the precise statement:

THEOREM 1. *To each congruence class* $\bmod p$ *of integers k admitting a solution of $x^2 - py^2 = k$ there exists a set of precisely p closed reflections on $\Gamma_p^*\backslash\mathcal{H}$. Different congruence classes have disjoint sets of reflections. Reflections within a set belonging to a single congruence class are conjugated into one another by $\Gamma_p^*\backslash\mathcal{H}$ conformal isometries given by the group generated by* $\begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}$. *All of these reflections are $\Gamma^0(p)$ conjugate to $L_0 := z \to p/\overline{z}$; and thus all of these reflections have fixed lines orthogonal to fixed lines of cusped reflections.*

Remark. Because $|k| < \sqrt{p} \Rightarrow k = (-1)^m Q_m$, and knowing the Q_m is equivalent to knowing $\mathrm{CF}(\sqrt{p})$, we see that the symmetries of the surface $\Gamma_p^*\backslash\mathcal{H}$ determine $\mathrm{CF}(\sqrt{p})$. This is one motivation for calling $\Gamma_p^*\backslash\mathcal{H}$ the $\sqrt{p}$ surface.

As stated above, [7] showed that if A-A-C is false at p, then there are p closed reflectors (called $\sigma_L = \pi(\pm\sqrt{p})$, $\sigma'_L, \sigma''_L, \ldots, \sigma_L^{(p)}$ below), of L_0 with cusped L_0-reversed reflectors orthogonal to them.

This certainly introduces so many new reflections on $\Gamma_p^*\backslash\mathcal{H}$ that hope of embedding it in $\mathbb{R}^3$ should be abandoned (see [6]); however, geometric

insight can be gained in such situations by passing to cleverly chosen finite covers.

Next, if we consider the shortest geodesic connections between these closed reflectors, we have

THEOREM 2. *If A-A-C is false at p, then the connections of the various $\sigma_L^{(n)}$ cannot intersect.*

As will be discussed in the next section of this Introduction, this offers both proof and algorithmic possibilities concerning A-A-C.

As will be shown below, Theorem 2 strongly suggested a study of when a closed geodesic on $\Gamma_p^*\backslash\mathcal{H}$ intersects a cusped reflector. This turned out to be completely determined by intermediate and plain convergents. Here is the precise result:

THEOREM 3. *Let $(\alpha,\overline{\alpha})$ be (project to) a closed geodesic on $\Gamma_p^*\backslash\mathcal{H}$. This geodesic will intersect an N_i reflector if and only if it intersects τ, a lift of that N_i reflector given by convergents and intermediate convergents of* $\mathrm{CF}(\alpha)$ *in accordance with exactly one of these four possibilities (assume $p_n/q_n < \alpha < p_{n+1}/q_{n+1}$, i.e., that n is odd*):

$$\text{(i) } \tau = \left(\frac{p_n + xp_{n+1}}{q_n + xq_{n+1}}, \frac{p_{n+1}}{q_{n+1}}\right),\ 0 \le x \le a_{n+2},$$

$$\text{(ii) } \tau = \left(\frac{p_{n+2}}{q_{n+2}}, \frac{xp_{n+2} + p_{n+1}}{xq_{n+2} + q_{n+1}}\right),\ 0 \le x \le a_{n+3},$$

$$\text{(iii) } \tau = \left(\frac{2p_n + xp_{n+1}}{2q_n + xq_{n+1}}, \frac{p_{n+1}}{q_{n+1}}\right),\ 1 \le x \le 2a_{n+2} - 1;\ x \text{ odd},$$

$$\text{(iv) } \tau = \left(\frac{p_{n+2}}{q_{n+2}}, \frac{xp_{n+2} + 2p_{n+1}}{xq_{n+2} + 2q_{n+1}}\right),\ 1 \le x \le 2a_{n+3} - 1;\ x \text{ odd},$$

where in every case p divides at least one of the four numerators or denominators. Note also that no orthogonality assumption is made as to the intersection.

1.3. New possibilities for A-A-C

• An explicit formula is given below for all lifts of σ'_L, but as is fully explained there, we cannot as yet effectively determine correct choices of two parameters (without actually computing y_0, which is of course the very thing we are trying to avoid). Now there are an infinite family of such correct choices; it would be beautiful if $\sigma'_L = (p^2 \pm \sqrt{p})/(p-1)$ were a correct lift; that is to say, that $(\alpha, p/\alpha) = (2 + p \pm \sqrt{(2+p)^2 + 4p})/2$ is always the shortest connection between σ_L and σ'_L. This is surely the "highest" connection in $\mathcal{H}$, thus motivating our hope.

If we could establish this for all p, or at least give an easy and effective decision algorithm, then we could either (a) show the algorithm given by Theorem 3 produces an N_0 reflector intersection with $(\alpha, p/\alpha)$—thus proving A-A-C; (b) run that algorithm on $(\alpha, p/\alpha)$ to test for the truth of A-A-C at p.

• We could simply run the algorithm given by Theorem 3 on $\sigma_L = (\pm\sqrt{p})$. According to [7], if this produced an N_i reflector intersection, for some $i \neq 0$, the A-A-C would be true at that p.

• The algorithm of Theorem 3 shows that N_0 reflector intersection with closed geodesics is determined by congruence conditions on pairs of "successive" intermediate and plain convergents. The congruence properties of successive plain convergents corresponding to a generic geodesic have been studied most illuminatingly by Moeckel [4]; the basic insight is that since the geodesic flow is known to be ergodic (and indeed much more), these congruence classes are randomly distributed (in appropriate sense).

What we are looking for here is congruences in some particular consecutive pair of convergents corresponding to a finite set of $p(p-1)/2$ specific geodesics—the shortest connections. Moeckel's work suggests that is tantamount to saying that at least one of these geodesics intersects a specific set in the tangent bundle on $\Gamma_p^*\backslash\mathcal{H}$. As many congruences are possible, this set may be large portion of that tangent bundle (a finite measure space). At the very least, this should yield a systematic reason to believe in the truth of A-A-C. It could also yield a proof that A-A-C is correct for an infinite (but indeterminate) set of primes.

2. Closed reflections on $\Gamma_p^*\backslash\mathcal{H}$ and complete quotients of $\sqrt{p}$.

The purpose of this section is to establish Theorem 1.

2.1. Closed reflections on $\Gamma_p^*\backslash\mathcal{H}$.

We begin by recalling (with slightly modified notation) the characterization of these reflections obtained in [7, Lemma A.4]:

Characterization. "[Every closed] reflection line may be lifted to the h-line in $\mathcal{H}$ running between $(\alpha p \pm \sqrt{p})/\gamma$. The action of the reflection is given by

$$z \to \begin{pmatrix} \alpha\sqrt{p} & -\beta\sqrt{p} \\ \gamma/\sqrt{p} & -\alpha\sqrt{p} \end{pmatrix} \overline{z}\,.$$

Here of course $\beta\gamma - \alpha^2 p = -1$ and $\alpha, \beta, \gamma \in \mathbb{Z}$."

We are going to associate reflections (note the plural) on $\Gamma_p^*\backslash\mathcal{H}$ with the k admitting solutions of $x^2 - py^2 = k$. If $|k| < p$, these arise from complete quotients of the continued fraction expansion of $\sqrt{p}$, and so we begin there.

2.1.1. $|k| < \sqrt{p}$. In what follows we will assume $(k,p) = 1$. If that fails, then $p \,|\, x$ and we have a Pell equation associated with a divisor of k. By iterating this process, we may assume without loss of generality that this divisor is prime to p.

Expand $\sqrt{p}$ in a regular continued fraction expansion $\mathrm{CF}(\sqrt{p}) = [a_0; a_1, a_2, \ldots]$ and as usual write

$$\frac{P_m + \sqrt{p}}{Q_m} = [a_m; a_{m+1}, a_{m+2}, \ldots]$$

for the mth complete quotient. It is classical that [3, p. 279]:

1. There are finitely many complete quotients (due to the periodicity of $\mathrm{CF}(\sqrt{p})$).
2. $Q_m < \sqrt{p}$.
3. $Q_m Q_{m-1} = p - P_m^2$.
4. $p_m^2 - q_m^2 = (-1)^m Q_m$, where p_m/q_m is the mth partial quotient.
5. Each k satisfying

$$x^2 - py^2 = k\,, \qquad |k| < \sqrt{p}$$

has the form $k = (-1)^m Q_m$ for some m.

Choose k with $(k,p) = 1$ satisfying the conditions of number 5 above; thus we have an m with $k = (-1)^m Q_m$. Let $\gamma = k$ in the Characterization. The problem is to determine α and β. Since $(p, Q_m) = 1$, we may find an *integer* $\equiv p^{-1} \bmod Q_m$. This is unique $\bmod Q_m$. Choose $\alpha \equiv P_m p^{-1} \bmod Q_m$. Since p^{-1} is unique $\bmod Q_m$, so is α. The role of the various choices of α within this constraint is discussed below.

We have

$$\alpha^2 p - 1 \equiv P_m^2 p^{-1} - 1 \equiv p^{-1}(P_m^2 - p) \equiv 0 \bmod Q_m$$

by number 3. This means that $\beta(-1)^m Q_m - \alpha^2 p = -1$ is soluble for β. This means there exist $\Gamma_p^* \backslash \mathcal{H}$ closed reflections with $(P_m^* p \pm \sqrt{p})/((-1)^m Q_m)$ as reflection lines where $P_m^* \equiv P_m \bmod Q_m$, $P_m^* = \alpha p$.

There are p different choices of $\alpha \bmod p$. This is because $\alpha_0 \equiv P_m p^{-1} \bmod Q_m$ may be chosen in $[0, Q_m)$ so that $\alpha_0, \alpha_0 + Q_m, \ldots, \alpha_0 + (p-1)Q_m$ are all different $\bmod p$.

This implies that different choices of α give different reflections. To see this, note that

$$(*)\quad \begin{pmatrix} \alpha\sqrt{p} & -\beta\sqrt{p} \\ \gamma/\sqrt{p} & -\alpha\sqrt{p} \end{pmatrix} \begin{pmatrix} \alpha'\sqrt{p} & -\beta'\sqrt{p} \\ \gamma'/\sqrt{p} & -\alpha'\sqrt{p} \end{pmatrix} \equiv_p \begin{pmatrix} -\beta\gamma' & -(\alpha\beta' - \alpha'\beta)p \\ \alpha'\gamma - \alpha\gamma' & -\beta'\gamma \end{pmatrix}.$$

(Here (and below) $\equiv_p$ means the upper right entry is considered mod p^2 while all other entries are mod p.)

Also, in our case $\gamma = \gamma' = k$, which forces $\beta \equiv \beta' \bmod p$ by the determinant condition on $\mathrm{SL}(2,\mathbb{R})$ matrices. This means RHS of $(*) \in \Gamma_p^* \Leftrightarrow \alpha \equiv \alpha' \bmod p$.

$(*)$ also implies different reflections with the same k are conjugate by an element of the subgroup $\mathbb{A}$ of $\Gamma^0(p)$ generated by $\begin{pmatrix} 1 & p \\ 0 & 1 \end{pmatrix}$.

Now we consider the question of whether two different Q_m's can give rise to the same reflection. (The congruences in this paragraph are all $\bmod p$.) Say $\gamma = (-1)^m Q_m$ and $\gamma' = (-1)^n Q_n$ give rise to the same reflection. Then

$$-1 \equiv \beta\gamma \equiv \beta'\gamma' \quad \text{and} \quad \beta'\gamma \equiv \beta\gamma' \equiv \pm 1\,.$$

The former follows from the determinant condition on reflections and the latter from the fact that RHS of $(*) \in \Gamma_p^*$. It is easy to see that choosing -1 in the second congruence forces $\gamma \equiv -\gamma'$ whereas choosing $+1$ forces $\beta \equiv \beta'$ and $\gamma \equiv \gamma'$. Recall now that all $0 < Q_j < \sqrt{p}$. Thus each congruence for the γ's forces $Q_m = Q_n$. This may happen [5], however the Pell equations involved are then identical (perhaps up to sign)—and this is no difference as $p \equiv 1 \bmod 4$.

2.1.2. $|k| > \sqrt{p}$**.** Again, we assume $(k,p) = 1$. We are going to associate reflections with k, just as we did in the previous case, by taking $k = \gamma$. The reduction (given for example in [3]) used to solve the Pell equation in the case $|k| > \sqrt{p}$ begins by finding U and k' with

$$U^2 - p = kk' \quad \text{and} \quad |k'| < |k|\,.$$

Now $(k,p) = (\gamma,p) = 1$ means we may find $p^{-1} \bmod \gamma$. Choose $\alpha \equiv Up^{-1} \bmod \gamma$. (Once again, α is unique $(\bmod \gamma)$.) This forces

$$\alpha^2 p \equiv U^2 p^{-1} \equiv 1 \bmod \gamma$$

and so we may solve $\beta\gamma - \alpha^2 p = -1$. Just as in the previous case we get p essentially different choices of α and the different reflections are conjugated into one another by elements of $\mathbb{A}$.

2.1.3. *Cusped reversed reflections on* $\Gamma_p^* \backslash \mathcal{H}$*.* We now turn to the question of when a reflection line $(\alpha p \pm \sqrt{p})/\gamma$ may be mapped onto $\pm\sqrt{p}$ by a conformal isometry. This in effect determines all cusped reversed reflectors as such an isometry will carry $(0,\infty)$, the imaginary axis which is a fixed line of the cusped reflection $N_0 := z \to -\overline{z}$, onto a reversed reflection line of another cusped reflection. An obvious necessary condition is $d^2 - pc^2 = \gamma$ be soluble as

$$g = \begin{pmatrix} a & bp \\ c & d \end{pmatrix} : \pm\sqrt{p} \to \frac{\alpha p \pm \sqrt{p}}{\gamma}$$

implies this. Recall that $L_0 := z \to p/\overline{z}$, and $(\alpha p \pm \sqrt{p})/\gamma$ is fixed by L_1. We wish to find $g \in \Gamma^0(p)$ with $g(\pm\sqrt{p}) = (\alpha p \pm \sqrt{p})/\gamma$.

That is, we wish to find $g \in \Gamma^0(p)$ with $L_1 g L_0 g^{-1} \in \Gamma_p^*$. Writing g in the notation of the previous paragraph and using the reflection matrices of the Characterization for the L's gives

$$L_1 g L_0 g^{-1} \equiv_p \begin{pmatrix} -\beta d^2 & [\alpha a^2 - \beta(ac-bd)]p \\ -\alpha d^2 - \gamma(bd-ac) & -\gamma a^2 \end{pmatrix}$$

and we want this $\in \Gamma_p^*$.

Note that as in the argument of the previous section, we get $a^2 \equiv \pm\beta$ and $d^2 \equiv \mp\gamma$ as necessary conditions. In addition, we need $p \mid -\alpha d^2 - \gamma(bd-ac)$ and $p \mid \alpha a^2 - \beta(ac-bd)$, which are both equivalent to

$$p \mid \pm\alpha - (ac-bd) \tag{1}$$

by $(\beta\gamma, p) = 1$ and the congruences on a^2 and d^2; the sign choice of these congruences determines the sign here.

We will show that a and b may be chosen so that equation (1) is satisfied. Fix c and d satisfying $d^2 - pc^2 = \gamma$. We may replace a by $a + xcp$ and b by $b + xd$ where $x \in \mathbb{Z}$ and we would still have $g \in \Gamma^0(p)$. Note that all such a's are congruent mod p, preserving $a^2 \equiv \pm\beta$.

The value of the RHS of equation (1) is now $\pm\alpha - ac + bd - x\gamma$ and by choosing x appropriately we may be sure that p divides this. Thus we have shown

THEOREM 4. *A necessary and sufficient condition for the reflector $(\alpha p \pm \sqrt{p})/\gamma$ to be a conformal isometric image of $\pm\sqrt{p}$ is that $d^2 - pc^2 = \gamma$ be soluble with $(c, d) = 1$.*

Remarks. (a) We call such closed reflectors *Pellian* because of the association with the equation.

(b) The Pellian reflectors are precisely those that have cusped reversed reflectors orthogonal to them, just as L_0 has $(\pm\sqrt{p})$ orthogonal to $(0, \infty)$ which is fixed by N_0.

(c) $\Gamma_p^* \backslash \mathcal{H}$ has closed reflectors that are not Pellian. Example ($p = 5$): Choose $\alpha = 1$, $\beta = \gamma = 2$. Then $\beta\gamma - \alpha^2 p = -1$ as required; and $(5 \pm \sqrt{5})/2$ is a closed reflection line. But $x^2 - 5y^2 = 2$ is not soluble as $(\frac{2}{5}) = -1$. So this reflector is not Pellian.

3. Connections between reflectors of L_0. This section discusses the number theoretic significance of the position of the shortest connections between closed reflector(s) of L_0 which intersect cusped reversed reflectors. (For this section we will henceforth drop the subscript and refer to this as $L : z \to p/\overline{z}$.) It will be recalled [7] that there is either one or p of these according as A-A-C is true or false at p. First, assume there are p such closed reflectors, $\sigma_L = \pi(\pm\sqrt{p})$, σ_L', σ_L'', ..., $\sigma_L^{(p)}$, recalling the notation of [7]. Consider the shortest geodesic connecting σ_L to another of these

reflectors, say σ'_L. There is no loss of generality in starting with σ_L since such reflectors may be mapped to one another by Γ_p^* conformal isometries (using the reversed reflectors orthogonal to them). We shall call this a *connection*. This connection must be orthogonal to both σ_L and σ'_L by local shortest path considerations. (Similar arguments show the connection is simple.) This orthogonality forces L to map this connection onto another one of equal length connecting σ_L to σ'_L; the two taken together form a closed geodesic—fixed, but not pointwise by L.

It is easy to see that this closed geodesic is in fact simple. For proceeding at constant speed in both directions from σ_L, at every moment we are at points paired by L. If we were to arrive at a self-intersection, it would have to be a fixed point of L and thus lie on a closed reflector. Minimality ensures we have arrived at σ'_L.

We will now prove that two such connections cannot intersect. We will show this yields a contradiction: Say α_1 connects σ_L and σ'_L and α_2 connects σ_L and σ''_L; and assume that they intersected at t (see Fig. 1):

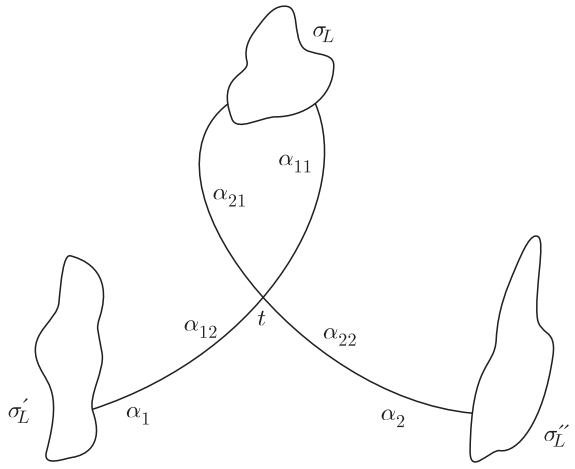

Fig. 1

Now t breaks α_i into two pieces, α_{i1} and α_{i2}, where the former leaves σ_L. Minimality implies $|\alpha_{21}| \geq |\alpha_{11}|$, where absolute value means length; this is by considering paths from σ_L and σ'_L. Symmetry now forces $|\alpha_{21}| = |\alpha_{11}|$ and now local arguments at t force the existence of a path between σ_L and σ'_L that is shorter than α_1. This completes the proof of Theorem 2.

We will now lift this picture to $\mathcal{H}$ under the assumption that A-A-C is false at p (see Fig. 2):

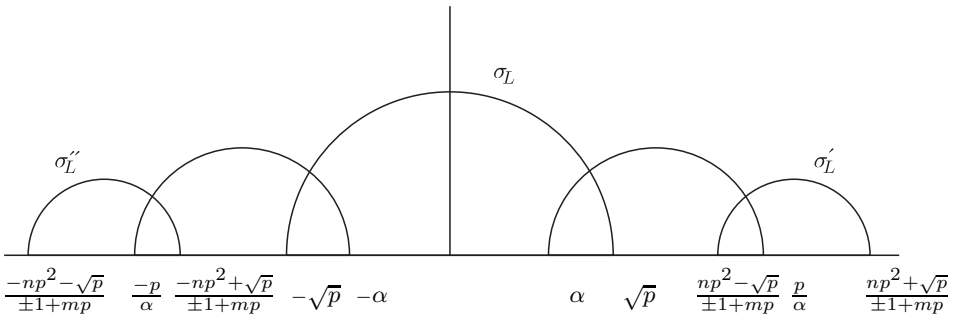

Fig. 2

In Figure 2, $\sigma_L = (\pm\sqrt{p})$ and σ'_L are presumed distinct reflectors of L, both reversed by cusped reflections. (Here and below we use $\sigma_L^{(n)}$ for both the reflectors on $\Gamma_p^*\backslash\mathcal{H}$ and their lifts to $\mathcal{H}$; this should cause no confusion.) Again, the assumption that $\sigma_L = (\pm\sqrt{p})$ is without loss of generality. Now there is a crucial difficulty here: If we select the *correct* lift of σ'_L (that is, if n and m are properly chosen in Figure 2) then the geodesic $(\alpha, p/\alpha)$ will be the shortest connection between σ_L and σ'_L. Otherwise it will just be some geodesic orthogonal to σ_L and σ'_L. And to make matters worse, it is easy to construct surfaces on which such arbitrary connections can indeed intersect. *Finding an efficient algorithm to determine the correct lift, without actually computing y_0, would offer an alternative test for the truth of A-A-C.*

Our next task is to develop some exact formuli for shortest connections and discuss the possibility of their intersecting. Note that $(\alpha, p/\alpha)$ is fixed *in* $\mathcal{H}$ by the product of the reflections through σ_L and σ'_L. (On $\Gamma_p^*\backslash\mathcal{H}$ both reflections are L, of course.) Thus α and p/α are fixed points of

$$\begin{pmatrix} 0 & \sqrt{p} \\ 1/\sqrt{p} & 0 \end{pmatrix}\begin{pmatrix} np\sqrt{p} & (\pm1+qp)\sqrt{p} \\ (\pm1+mp)/\sqrt{p} & -np\sqrt{p} \end{pmatrix}=\begin{pmatrix} \pm1+mp & -np^2 \\ np & \pm1+qp \end{pmatrix}.$$

Here of course $(\pm1+mp)(\pm1+qp)+n^2p^3=1$. This gives

$$\alpha = \frac{(m-q)+\sqrt{(m-q)^2+4n^2p}}{2n}\,.$$

As we have said, if m and n (and thus q) are properly chosen this is indeed a connection.

We now wish to consider the action of N_0 in order to obtain another connection. Recall that N_0 and L commute as they have orthogonal reflectors. This means that

$$N_0\left(\left(\frac{np^2\pm\sqrt{p}}{\pm1+mp}\right)\right) = \left(\frac{-np^2\pm\sqrt{p}}{\pm1+mp}\right)$$

is also an L reflector orthogonal to a cusped reflector. Also $N_0((\alpha, p/\alpha)) = (-\alpha, -p/\alpha)$ is the connection between $\left(\frac{-np^2\pm\sqrt{p}}{\pm1+mp}\right)$ and $(\pm\sqrt{p})$.

These connections would be identical if $\left(\frac{np^2\pm\sqrt{p}}{\pm 1+mp}\right)$ and $\left(\frac{-np^2\pm\sqrt{p}}{\pm 1+mp}\right)$ were the same geodesic δ on $\Gamma_p^*\backslash\mathcal{H}$. But this is not possible because they are identified by N_0. Specifically: (i) δ could not be fixed by N_0 because it is closed, (ii) δ could not be not reversed by N_0 because we are assuming that A-A-C is false and Lemma D2 of [7] applies, that is, σ_L is the unique reflector reversed by N_0, (iii) N_0 could not act as a glide-reflection along δ as such geodesics have the form

$$\frac{-(a+d)\pm\sqrt{(a+d)^2-4bcp^3}}{2cp} \quad \text{where} \quad \begin{pmatrix} a & bp^2 \\ cp & d \end{pmatrix} \in \Gamma_p^*$$

and $(np^2\pm\sqrt{p})/(\pm 1+mp)$ cannot be of this form. At this point, we know that if $(\alpha, p/\alpha)$ intersects $(-\alpha,-p/\alpha)$ we would have a contradiction to Theorem 2 and A-A-C would be established at p. That is, an intersection is sufficient to establish A-A-C. Perhaps the easiest way to establish the intersection would be to show that $(\alpha, p/\alpha)$ intersects an N_0 reflector. (For we have just shown that this intersection would be on $(-\alpha,-p/\alpha)$ also.)

(In the next section we shall determine precisely which geodesics have an N_0 reflector intersection in terms of intermediate convergents. And indeed this method works for all cusped reflections N_i on the $\Gamma_p^*\backslash\mathcal{H}$.)

We close this section with a conjecture concerning the necessity of such an intersection for A-A-C.

CONJECTURE. *If A-A-C is true then $(\alpha, p/\alpha)$ intersects $(-\alpha,-p/\alpha)$. Moreover, there is such an intersection caused by intersection with an N_0 reflector.*

4. Convergents and cusped reflector intersections. The purpose of this section is to prove Theorem 3.

We begin with the N_0 case. Recall that an N_0 reflector has the form $(a/pc, p^2b/d)$ with $ad-p^3bc=\pm 1$ [7, Lemma B5]. We have $p_n/q_n<\alpha<p_{n+1}/q_{n+1}$ and $p_{n+1}q_n-q_{n+1}p_n=\pm 1$. If $p\,|\,q_n$ and $p^2\,|\,p_{n+1}$, we are done. (Here $|\alpha-p/\alpha|$ is assumed large; i.e. greater than one.) We are going to modify this idea using intermediate convergents—some of what follows immediately is standard computation with intermediate convergents.

We want

$$\frac{sp_n+rp_{n+1}}{sq_n+rq_{n+1}}<\alpha<\frac{tp_n+up_{n+1}}{tq_n+uq_{n+1}}\,;$$
$$p\,|\,sq_n+rq_{n+1}\,, \quad p^2\,|\,tp_n+up_{n+1}\,; \quad r/s\,,\ t/u>0\,,$$

with

$$(sp_n+rp_{n+1}))(tq_n+uq_{n+1})-(sq_n+rq_{n+1})(tp_n+up_{n+1})=\pm 1,\,\pm 2\,.$$

(This last may be written:

$$\det \begin{pmatrix} s & r \\ t & u \end{pmatrix} \begin{pmatrix} p_n & q_n \\ p_{n+1} & q_{n+1} \end{pmatrix} = \pm 1, \pm 2$$

and thus $su - rt = \pm 1, \pm 2$ is necessary and sufficient.) We begin with the case of $+1$.

Write $x = r/s$ and consider $(p_n + xp_{n+1})/(q_n + xq_{n+1}) = \alpha$. Clearly $x \notin \mathbb{Z}$, and as a function of x, clearly $(p_n + xp_{n+1})/(q_n + xq_{n+1})$ increases with x. This means there is a unique integer b such that

$$\frac{p_n + bp_{n+1}}{q_n + bq_{n+1}} < \alpha < \frac{p_n + (b+1)p_{n+1}}{q_n + (b+1)q_{n+1}}$$

and it is now a simple matter using standard convergent facts to show that $b = a_{n+2}$. We will continue to use b as a notational convenience.

We have $r/s \leq b$ or $r \leq bs$, and similarly $t(b+1) \leq u$. Recall that $u = (1 + rt)/s \leq 1/s + bt$. Thus $bt + t \leq u \leq bt + 1/s$. This forces $s = t = 1$ with $u = b + 1$ and $r = b$; or $t = 0$ with $s = u = 1$ and $0 \leq r \leq b$. The first case is really that of a convergent p_{n+2}/q_{n+2} and its first intermediate convergent toward p_{n+1}/q_{n+1}. The second is

$$\left(\frac{p_n + ip_{n+1}}{q_n + iq_{n+1}}, \frac{p_{n+1}}{q_{n+1}} \right) \quad \text{for } 0 \leq i \leq b\,.$$

We can now exhaust the $\det = +1$ possibilities of N_0 reflectors surrounding α that come from intermediate convergents. In the following *test* means check numerators and denominators of the intervals for the required congruence conditions.

ALGORITHM. *Begin with* $(\frac{p_n}{q_n}, \frac{p_{n+1}}{q_{n+1}})$. *Test* $(\frac{p_n + ip_{n+1}}{q_n + iq_{n+1}}, \frac{p_{n+1}}{q_{n+1}})$ *for* $i \leq a_{n+2}$. (*The last test is* $(\frac{p_{n+2}}{q_{n+2}}, \frac{p_{n+1}}{q_{n+1}})$). *Proceed to test* $(\frac{p_{n+2}}{q_{n+2}}, \frac{ip_{n+2} + p_{n+1}}{iq_{n+2} + q_{n+1}})$ *for* $i \leq a_{n+3}$. (*The last test is* $(\frac{p_{n+2}}{q_{n+2}}, \frac{p_{n+3}}{q_{n+3}})$). *Continue in this way.*

The $\det = +2$ *case is similar*: *Begin with* $(\frac{2p_n + ip_{n+1}}{2q_n + iq_{n+1}}, \frac{p_{n+1}}{q_{n+1}})$. (*If* i *is even, this has already occurred in the* $\det = +1$ *case, so we restrict to* i *odd.*) *We test for* $i \leq 2a_{n+2} - 1$. (*The last test is* $(\frac{2p_{n+2} - p_{n+1}}{2q_{n+2} - q_{n+1}}, \frac{p_{n+1}}{q_{n+1}})$; *note that* $i = 2a_{n+2} + 1$ *gives* $(\frac{2p_{n+2} + p_{n+1}}{2q_{n+2} + q_{n+1}}, \frac{p_{n+1}}{q_{n+1}})$ *and this entire interval is to the right of* α.) *Proceed to test* $(\frac{p_{n+2}}{q_{n+2}}, \frac{ip_{n+2} + 2p_{n+1}}{iq_{n+2} + 2q_{n+1}})$, *for odd* $i \leq 2a_{n+3} - 1$. *Continue in this way.*

The $\det = -1, -2$ cases simply reverse the role of n and $n + 1$; i.e. assume that n is even. Also, the generalization from N_0 to N_i, as per [7, Lemmas A3 and B1], amounts to checking different congruence conditions as we test. Alternatively we could (effectively) conjugate N_i into N_0 with a $\Gamma_p^* \backslash \mathcal{H}$ conformal isometry ϕ and then test $\phi(\alpha)$ for intersection by an N_0 reflector.

Last, we will show that all N_i reflectors crossing $(\alpha, \overline{\alpha})$ arise from intermediate and plain convergents. Indeed, if $|ad-bc| = 1$, we have $|a/c-b/d| = 1/(cd)$. Without loss of generality assume that $|\alpha - a/c| > |\alpha - b/d|$, and $a/c < \alpha < b/d$. If $c > d$ then $|\alpha - b/d| \leq 1/(2cd) \leq 1/(2d^2)$ and b/d is a convergent, say p_{n+1}/q_{n+1}. This forces $a = p_n + xp_{n+1}$ and $c = q_n + xq_{n+1}$. (If $p_n/q_n < a/c$, we have $x \leq a_{n+2}$; we will deal with this assumption at the end of the proof.)

We are reduced to the case $d > c$, $|\alpha - a/c| > |\alpha - b/d|$. This forces $|\alpha - a/c| < 1/c^2$ and $a = p_n + xp_{n+1}$, $c = q_n + xq_{n+1}$ where again $p_n/q_n < a/c$ gives $0 \leq x \leq a_{n+2}$. Just as before we get $b = p_{n+1} + y(p_n + xp_{n+1})\,, d = q_{n+1} + y(q_n + xq_{n+1})$. Since $\alpha < b/d$, we have

$$\frac{xy+1}{y} = x + \frac{1}{y} \geq a_{n+2} \geq x.$$

This resolves to the two cases $y = 1$, $x = a_{n+2} - 1$ or $a_{n+3} \geq y > 1$, $x = a_{n+2}$; the latter of which is part of the p_{n+2}/q_{n+2} intermediate convergent case. The case $|ad - bc| = 2$ proceeds similarly.

The assumption that $p_n/q_n < a/c < \alpha$ is without loss of generality. This is because a lift of any N_i reflector intersecting $(\alpha, \overline{\alpha})$ may be mapped (by a power of the primitive Γ_p^* hyperbolic fixing $(\alpha, \overline{\alpha})$) onto a lift of the same N_i reflector with $p_n/q_n < a/c < p_{n+2}/q_{n+2} < \alpha$ for some n (not uniquely, of course). This completes the proof of Theorem 3.

References

[1] M. Akbas and D. Singerman, *Symmetries of modular surfaces*, preprint.

[2] N. C. Ankeny, E. Artin and S. Chowla, *The class number of real quadratic fields*, Ann. of Math. (2) 56 (1952), 479–493.

[3] L. K. Hua, *Introduction to Number Theory*, Springer, New York 1981.

[4] R. Moeckel, *Geodesics on modular surfaces and continued fractions*, Ergodic Theory Dynamical Systems 2 (1982), 69–83.

[5] R. Mollin and H. C. Williams, *Class number one for real quadratic fields, continued fractions, and reduced ideals*, in: Canadian Number Theory Association Conference Proceedings (Banff, 1988), R. Mollin (ed.), W. de Gruyter, Berlin 1990, 417–425.

[6] R. Ruedy, *Symmetric embeddings of Riemann surfaces*, in: Discontinuous Groups and Riemann Surfaces, Proc. Conf. (Univ. Maryland, College Park, Md., 1973), Ann. of Math. Stud. 79, Princeton Univ. Press, Princeton, N.J., 1974, 409–418.

[7] M. Sheingorn, *Hyperbolic reflections on Pell's equation*, J. Number Theory 33 (1989), 267–285.

DEPARTMENT OF MATHEMATICS
CUNY, BARUCH COLLEGE
NEW YORK, NEW YORK 10010
U.S.A.

Received on 17.10.1991 (2186)