References

[1] A. Oppenheim, *On the representation of real numbers by products of ration numbers*, Quart. J. Math. Oxford 4 (1953), pp. 303–307.
[2] O. Perron, *Irrationalzahlen*, Chelsea, N. Y., 1948.

UNIVERSITY OF GHANA
Legon, Ghana

Received on 16. 1. 1970

ACTA ARITHMETICA
XVIII (1971)

Representation of Markoff's binary quadratic forms by geodesics on a perforated torus

by

HARVEY COHN (Tucson, Arizona)

In memory of Harold Davenport

1. Introduction. One of Harold Davenport's most remarkable contributions was a succession of papers (notably [4] and [5]) on the minima of the product of three ternary homogeneous linear forms (compare Mordell [11]). Davenport showed that (with unit determinant) the two largest minima, $1/7$ and $1/9$, are discrete. No further minima have been established since then.

One of the reasons that this problem is so intriguing and challenging is the comparison one naturally makes with the Markoff theory of binary (indefinite) quadratic forms (see [10], [6], [2]). The Markoff theory represents a state of perfection at the fringes of utter chaos! A discrete, convergent sequence of minima exists with a limit point $(1/3)$ below which the spectrum of minima varies locally from continuous to discrete ([9], [12]). The original theory depended heavily on continued fractions, although a revision of Frobenius [7] made the theory depend more on chains of reduced forms. A paper of the author [3] used as a substitute tool some algebraic (matrix) identities which, in principle, are less specialized than continued fractions.

We now return to our earlier approach [3] in the hope that additional insight might be gained in understanding the discrete nature of the minima by an exploration of the geometric aspects of the Markoff forms. We interpret these forms in terms of closed geodesics of preassigned homology type on a perforated torus. It is possible, specifically, to gain a better understanding of some of the "fringe" behavior at the limit point of the discrete set of minima.

2. Rational Markoff forms. We briefly summarize the classical theory. Let [10]

(1) $$Q(x, y) = ax^2+bxy+cy^2, \quad d = b^2-4ac > 0$$

be a quadratic form with integral coefficients and irrational roots ξ, η (i.e., $Q(x, y) = a(x-\xi y)(x-\eta y)$). Let us define the minima

$$m(Q) = \inf |Q(x, y)|/d^{1/2} \tag{2}$$

taken over integer-pairs $(x, y) \neq (0, 0)$. Then consider the sequence of Markoff triples $M^{(j)} = (M_1^{(j)}, M_2^{(j)}, M_3^{(j)})$ given by the positive solutions of $(M_1^{(j)})^2+(M_2^{(j)})^2+(M_3^{(j)})^2 = 3M_1^{(j)} M_2^{(j)} M_3^{(j)}$ arranged so that every solution (Markoff number) occurs once in order $M_1^{(1)} < M_1^{(2)} < M_1^{(3)} < \ldots$ We construct the forms

$$Q_j(x, y) = a_j x^2+b_j xy+c_j y^2 \tag{3}$$

where $a_j = M_1^{(j)}$, $b_j = (2u_j-3M_1^{(1)})$, $c_j = (v_j-3p_j)$, u_j is the least positive residue of $\pm M_2^{(j)}/M_3^{(j)}$ (modulo $M_1^{(j)}$) and $v_j = (1+u_j^2)/M_1^{(j)}$. Here $m(Q_j) = (9-(2/M_1^{(j)})^2)^{-1/2} > 1/3$; and the minima $m(Q_j)$ are discrete and have $1/3$ as a limit point. The first three cases are

$$\boldsymbol{M} = (1, 1, 1), \quad Q_1(x, y) = x^2-3xy+y^2, \quad m(Q_1) = 1/5^{1/2};$$

then

$$\boldsymbol{M} = (2, 1, 1), \quad Q_2(x, y) = 2x^2-4xy-2y^2, \quad m(Q_2) = 1/8^{1/2};$$

and finally

$$\boldsymbol{M} = (5, 2, 1), \quad Q_3(x, y) = 5x^2-11xy-5y^2, \quad m(Q_3) = 5/221^{1/2}.$$

Markoff's famous result is that if

$$m(Q) > 1/3, \tag{4}$$

then Q is a rational multiple of a form equivalent to a Q_i.

In the alternate derivation [3], the sequence of Markoff triples were seen to be one-third the traces of matrix triples serving as generating elements of an automorphic group with fundamental domain of genus one. In the interpretation (as geodesics) undertaken here we shall require some of the continued fraction theory (see [6]), which must be summarized for reference.

3. Markoff symmetry property. Let ξ and η be roots of the Markoff form (assumed reduced) so

$$\xi > 1 > -\eta > 0. \tag{5}$$

Then ξ and $-\eta$ are given by continued fractions in the usual notation

$$\xi = a_0+1/(a_1+1/(a_2+\ldots = \{a_0, a_1, a_2, \ldots\}, \tag{6a}$$

$$-\eta = 1/(a_{-1}+1/(a_{-2}+\ldots = \{0, a_{-1}, a_{-2}, \ldots\}, \tag{6b}$$

with a_i positive integers. Using the obvious notation

$$\{-\eta, \xi\} = \{\ldots, a_{-2}, a_{-1}, a_0, a_1, a_2, \ldots\}, \tag{7}$$

we obtain a periodic sequence (Markoff period) consisting of integer pairs 1, 1 or 2, 2 satisfying the following symmetry property:

Write the periodic sequence as

$$\{\ldots, (1, 1)_{r_{-1}}, 2, 2, (1, 1)_{r_0}, 2, 2, (1, 1)_{r_1}, 2, 2, \ldots\} = [\ldots, r_{r_{-1}}, r_0, r_1, \ldots] = \boldsymbol{R}. \tag{8}$$

(Clearly this omits $\{\ldots, 1, 1, 1, 1, \ldots\}$ which we might designate as $[\infty] = \boldsymbol{R}$; but $\{\ldots, 2, 2, 2, 2, \ldots\}$ would be merely $[\ldots, 0, 0, \ldots] = \boldsymbol{R}$.) Then the Markoff symmetry property is that

(i) the items of sequence $\boldsymbol{R}$ have at most two values of the form $r, r+1$ (or $r, r-1$), and

(ii) if $r_j \neq r_{j+1}$ we go to the right and left a distance h to form $d_h = r_{j-h}-r_{j+1+h}$. Then the first $h > 0$ where $d_h \neq 0$ is one where d_0 and d_h disagree in sign.

The symmetry property is not disturbed if we add or subtract a constant to each item. Also the symmetry property is vacuously satisfied if all r_j are equal. This can serve as an effective criterion of symmetry owing to Markoff's *derivation condition.* If we write $\boldsymbol{R}$ as

$$\boldsymbol{R} = [\ldots, (r\pm 1)_{s_{-1}}, r, (r\pm 1)_{s_0}, r, \ldots] \tag{9}$$

then we obtain a *derived sequence*

$$\boldsymbol{R}' = \boldsymbol{S} = [\ldots, s_{-1}, s_0, s_1, s_2, \ldots] \tag{10}$$

which has the Markoff symmetry property if and only if $\boldsymbol{R}$ has the property. It is not hard to show that any Markoff period can be reduced to a trivial one (all the same value) by derivation.

4. Juxtaposition of periods. It is important for latter purposes to construct Markoff periods by juxtaposition of shorter periods rather than derivation (which serves us only as a tool for the proofs). Certainly not every juxtaposition of $(1, 1)$ and $(2, 2)$ is a Markoff period (e.g., $\{\ldots, 1, 1, 2, 2, 1, 1, 2, 2, 2, 2, 2, 2, \ldots\}$ fails to satisfy the symmetry condition). To begin with, we need this theorem, essentially due to Markoff [10]:

Let (u, v) be two relatively prime positive (or zero) integers, where u denotes the number of pairs $(2, 2)$ and v denotes the number of pairs $(1, 1)$. Then all Markoff numbers correspond to such pairs uniquely with $M(u, v)$ formed by the following rules: First,

$$M(0, 1) = 1, \quad M(1, 0) = 2, \quad M(1, 1) = 5 \tag{11a}$$

and the general Markoff triple $M_i = M(u_i, v_i)$ $(i = 1, 2, 3)$ correspond to a general triple of (u_i, v_i) of which any two are unimodular, $u_i v_j - u_j v$ $= \pm 1$ $(i \neq j)$. The chain rule of formation is that if

(11b) $(M_1, M_2, M_3) = \big(M(u_1, v_1), M(u_2, v_2), M(u_3, v_3)\big)$

then (for $3M_1M_2 > M_3$)

(11c) $(M_1, M_2, 3M_1M_2 - M_3)$

$$= \big(M(u_1, v_1), M(u_2, v_2), M(u_1+u_2, v_1+v_2)\big).$$

(The same rule holds, of course, with any permutation of indices.)

This result, actually, follows very easily from the basis constructio given in [3]. Note that to start the induction,

(12a) $(u, v) = (0, 1)$,

$$\boldsymbol{R} = [\infty], \quad \{-\eta, \xi\} = \{\ldots, 1, 1, \ldots\}, \qquad Q = Q_1 \ (M = 1),$$

(12b) $(u, v) = (1, 0)$,

$$\boldsymbol{R} = [0], \quad \{-\eta, \xi\} = \{\ldots, 2, 2, \ldots\}, \qquad Q = Q_2 \ (M = 2),$$

(12c) $(u, v) = (1, 1)$,

$$\boldsymbol{R} = [0, 1], \quad \{-\eta, \xi\} = \{\ldots, 1, 1, 2, 2, \ldots\}, \qquad Q = Q_3 \ (M = 5).$$

Thus $M(1, 2) = 13$, $M(2, 1) = 29$, $M(1, 3) = 34$, $M(1, 4) = 89$, $M(3, 1$ $= 169$, $M(2, 3) = 194$, $M(1, 5) = 233$, $M(3, 2) = 433$, $M(1, 6) = 610$ $M(4, 1) = 985$ (which accounts for all $M < 1000$).

We now consider the general Markoff period

(13a) $\boldsymbol{P}(u, v) = \{\ldots, 2, 2, \ldots, 1, 1, \ldots\}, \quad (u, v) = 1, \ u \geqslant 0, \ v \geqslant 0,$

and divide all such periods into two types (overlapping only whe $u = v = 1$). Here $\boldsymbol{P}(u, v)$ is of type 1 or 2 depending on whether ther are more ones $(v \geqslant u)$ or more twos $(u \geqslant v)$. For every $\boldsymbol{P}(u, v)$ of (say type 2 we have two *normalized forms of the period* (called *upper* and *lower*) namely

(13b) $\boldsymbol{P}_+(u, v) = \{1, 1, \ldots, 2, 2\}$; $\quad \boldsymbol{P}_-(u, v) = \{2, 2, \ldots, 2, 2\}$

(except $\boldsymbol{P}_+(1, 1) = \{1, 1, 2, 2\} = \boldsymbol{P}_-(1, 1)$), with the property tha various periods $\boldsymbol{P}_\pm(u, v)$ of type 2 can be placed in juxtaposition by th following rules of the continued fraction:

Let the convergents of $\gamma = \{g_0, g_1, \ldots\}$ *be* u_n/v_n $(g_0 > 0)$, *so*

(14a) $(u_{n+2}, v_{n+2}) = g_{n+2}(u_{n+1}, v_{n+1}) + (u_n, v_n)$.

Then for n *odd* (i.e., $u_{n+1}/v_{n+1} < u_n/v_n$),

(14b) $\boldsymbol{P}_-(u_{n+2}, v_{n+2}) = \boldsymbol{P}_-(u_n, v_n) + g_{n+2}\boldsymbol{P}_-(u_{n+1}, v_{n+1})$,

(14c) $\boldsymbol{P}_+(u_{n+2}, v_{n+2}) = g_{n+2}\boldsymbol{P}_+(u_{n+1}, v_{n+1}) + \boldsymbol{P}_+(u_n, v_n)$;

while for n *even* (i.e., $u_{n+1}/v_{n+1} > u_n/v_n$),

(14d) $\boldsymbol{P}_-(u_{n+2}, v_{n+2}) = g_{n+2}\boldsymbol{P}_-(u_{n+1}, v_{n+1}) + \boldsymbol{P}_-(u_n, v_n)$,

(14e) $\boldsymbol{P}_+(u_{n+2}, v_{n+2}) = \boldsymbol{P}_+(u_n, v_n) + g_{n+2}\boldsymbol{P}_+(u_{n+1}, v_{n+1})$,

with similar relations for type 1. (*Note the noncommutative juxtaposition of words in* (14b) *to* (14e).)

5. Polygonal model of period. The proof of these assertions consists of an examination of the sequence $\boldsymbol{R}$ (see (8)) corresponding to a general period $\boldsymbol{P}$ (say) of type 2. First note the $\boldsymbol{R}$ must consist of only zeros and ones. (For if there are r values of $(k+1)$ and s values of (k) in a period of $p = r+s$ in $\boldsymbol{R}$, then $u = r+s$ values of 2, 2 occur in $\boldsymbol{P}$ and $v = ks+$ $+(k+1)r$ values of 1, 1 occur in $\boldsymbol{P}$, which requires that $k = 0$ in order that $u \geqslant v$). Now we suppose $\boldsymbol{R}$ to be a period of $p = (r+s)$ terms with r ones and s zeros, where $(r, s) = 1$. We define an *upper* and *lower derived period* $\boldsymbol{R}_+$ (for $\boldsymbol{P}_+$) and $\boldsymbol{R}_-$ (for $\boldsymbol{P}_-$) as follows: Define

(15) $a_t = [tr/s] - [(t-1)r/s], \quad A_t = -[-tr/s] + [-(t-1)r/s].$

(Note $-[-x]$ is the next "large or equal" integer, while $[x]$ is the next "smaller or equal"). Then define (for $1 \leqslant j \leqslant s$)

(16a) $v_j = 0, 1, \ldots, 1$ (with a_j elements 1 in juxtaposition),

(16b) $V_j = 0, 1, \ldots, 1$ (with A_j elements 1 in juxtaposition).

Then the normalized upper and lower derived periods are

(17a) $$\boldsymbol{R}_+ = [V_1, \ldots, V_s],$$

(17b) $$\boldsymbol{R}_- = [v_1, \ldots, v_s].$$

The upper and lower Markoff periods (of type 2) are uniquely determined from them $(u = r+s > v = r)$. Through an accident of notation, the "upper" polygons for r/s correspond to "lower" polygons for u/v.

Before proving that $\boldsymbol{R}_+$ and $\boldsymbol{R}_-$ have the Markoff symmetry property (as well as the juxtaposition property) it is well to interpret them geometrically as in Figure 1. Here $\boldsymbol{R}_-$ (and $\boldsymbol{R}_+$) are seen to represent the lower (and upper) polygon $\mathscr{P}_-$ and $\mathscr{P}_+$ approximating the diagonal from $(0, 0)$ to (r, s) so that *no integral lattice point lies between the polygon and the diagonal.* In a sense, $\mathscr{P}_+$ and $\mathscr{P}_-$ are the "straightest" polygonal approximations to the diagonal; later on we shall see they determine the paths of geodesics!

The construction is illustrated for $r = 5$, $s = 3$ so that

$$\boldsymbol{R}_- = [0, 1, 0, 1, 1, 0, 1, 1], \qquad \boldsymbol{R}_+ = [1, 1, 0, 1, 1, 0, 1, 0],$$

and by (8) we retrieve

$$\boldsymbol{P}_-(8, 5) = \{2, 2, 1, 1, 2, 2, 2, 2, 1, 1, 2, 2, 1, 1, 2, 2, 2, 2, 1, 1, \\ 2, 2, 1, 1, 2, 2\},$$

$$\boldsymbol{P}_+(8, 5) = \{1, 1, 2, 2, 1, 1, 2, 2, 2, 2, 1, 1, 2, 2, 1, 1, 2, 2, 2, 2, \\ 1, 1, 2, 2, 2, 2\}.$$

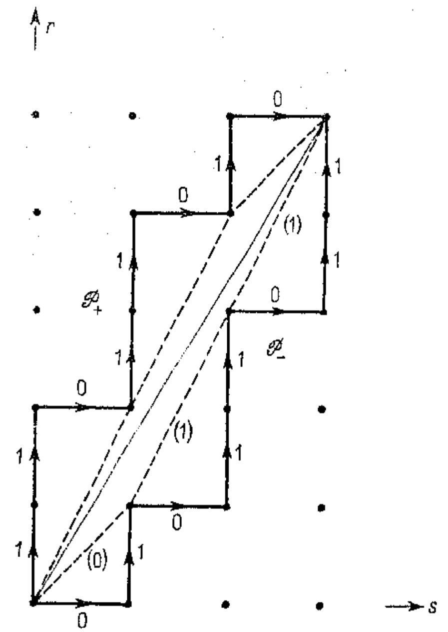

Fig. 1. Upper and lower polygons $\mathscr{P}_+$ and $\mathscr{P}_-$. They represent the normalized uppe and lower derived periods $\boldsymbol{R}_+$ and $\boldsymbol{R}_-$ for $(r, s) = (5, 3)$

The corresponding Markoff number is 48, 928, 105 from the rule (11c)

To see that (say) $\boldsymbol{R}_-$ satisfies the Markoff symmetry condition write it as $\boldsymbol{R}_- = [0, 1_{a_1}, 0, 1_{a_2}, \ldots, 0, 1_{a_s}]$, so its derived period is th shorter expression $\boldsymbol{R}'_- = [a_1, a_2, \ldots, a_s]$, with each $a_j = [r/s]$ or $1 + [r/s]$ If we subtract $[r/s]$ from each item of $\boldsymbol{R}'_-$, we get a set of r' ones and s zeros where (for $k = [r/s]$),

$$r' = r - sk, \qquad s' = sk + s - r. \tag{18}$$

Now we can see this new set of $r' + s'$ zeros and ones is the $\boldsymbol{R}_-$ of (r', s'). It is easiest to see this by the fact that (18) is an affine unimodular transformation (with k paradoxically imagined to be a constant!) And the property of the lower and upper polygons are unimodular-affine invariant. (In Figure 1, the direction of the new coördinates is shown by the dotted lines. Here $(r', s') = (1, 2)$, e.g., $\boldsymbol{R}'_- = [1, 2, 2]$, and subtracting $[r/s] = 1$, we get $[0, 1, 1]$, illustrated as the dotted lower polygon in the new coordinates.)

The juxtaposition relations (14b) to (14e) now follow almost immediately from the H. J. S. Smith geometric construction of the continued fraction ([9], p. 38). The essence of the construction is that the vector sum (14a) produces the resultant (u_{n+2}, v_{n+2}) as a diagonal, but no lattice points occur between the addends and the diagonal. (In Figure 1, $\gamma = \{1, 1, 2, \ldots\}$, so $u_0/v_0 = 1$, $u_1/v_1 = 2$, $u_2/v_2 = 5/3$.) Thus the "upper and lower polygons" are preserved by the juxtaposition.

6. Representation of forms by geodesics in the upper half plane. We begin with the traditional Klein–Poincaré non-Euclidean model of the upper half z plane with the metric

$$ds = |dz|/(\mathrm{Im}\, z). \tag{19}$$

Thus if ξ and η are any real numbers, then we can represent the geodesics as $G(\eta, \xi)$ the upper semicircle from η to ξ with center on the real axis. This semicircle can be represented in the fundamental domain $\boldsymbol{D}_0$ of the modular group Γ_1, namely

$$\boldsymbol{D}_0\colon\ |z| > 1, \quad |\mathrm{Re}\, z| \leqslant 1/2 \tag{20}$$

by integral fractional unimodular transformations [8]. We shall find it more convenient to consider the image in the union of integral translates of $\boldsymbol{D}_0$, namely

$$\boldsymbol{D}_\infty\colon\ \bigcap |z - n| \geqslant 1 \quad (\mathrm{Im}\, z > 0), \tag{21}$$

for all integers n. Now the geodesic $G(\eta, \xi)$ can be represented by use of an additional symmetry operation $z \to -\bar{z}$ (together with the modular group). Then the geodesics continually "bounce" along the 60-degree arcs forming the base of $\boldsymbol{D}_\infty$ (see Figure 2).

Thus the (unoriented) geodesic in $\boldsymbol{D}_\infty$ for roots η, ξ corresponds to the equivalence class of quadratic forms $eQ(x, y)$ $(e \neq 0)$, with rational coefficients. They are closed geodesics under translation by even integers (compare [8]).

On a geodesic arc we note the

$$\sup(\mathrm{Im}\, z) = \big(2m(Q)\big)^{-1}, \tag{22}$$

because the (ordinary) radius of a geodesic is $|(\xi-\eta)/2| = d^{1/2}/(2a)$, where the first coefficient, a, can take the minimum of the form.

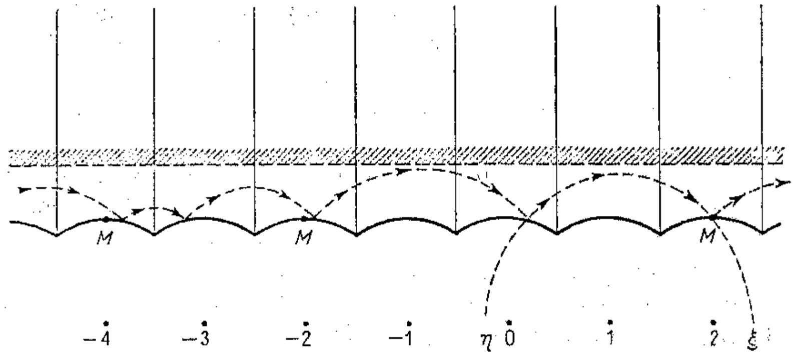

Fig. 2. Path of geodesic, represented in $\boldsymbol{D}_\infty$, corresponding to $\{\ldots, 1, 1, 2, 2, \ldots\}$.

Thus the Markoff forms are characterized by the translationally closed geodesics which remain below the horizontal line $\operatorname{Im} z = 3/2$ (see the dotted line in Figure 2). We ignore the trivial case of a rational root ξ or η, because there it is seen that the geodesic will go straight up to ∞ somewhere.

The relation between the "jumps" of the geodesic in $\boldsymbol{D}_\infty$ and the denominators of the continued fraction is a generally very annoying one, but here the nature of the Markoff form is more than adequate to simplify matters.

Let $Q(x, y)$ be a quadratic form whose period contains only ones and twos (not necessarily paired). Then the geodesic connecting the roots $G(\eta, \xi)$ enters $\boldsymbol{D}_\infty$ (starting from η) in the interval $|\operatorname{Re} z| < 1/2$ and leaves $\boldsymbol{D}_\infty$ (continuing to ξ) in the interval $|\operatorname{Re} z - a_t| < 1/2$, where in the usual notation

(23) $$\xi = \{a_t, a_{t+1}, \ldots\}, \qquad -\eta = \{0, a_{t-1}, \ldots\} \qquad (a_j \geqslant 1).$$

Thus the "jumps" of such a geodesic exactly display the denominators of the continued fraction.

Let us consider, for proof, the entry (from η). We are concerned with certain distance inequalities making the radius $(\xi-\eta)/2$ lie between the distance from the center $(\xi+\eta)/2$ and the end-points of the interval $\pm 1/2 + i3^{1/2}/2$ (see Figure 2). Thus

(23a) $$\left(\tfrac{1}{2}(\xi+\eta)-\tfrac{1}{2}\right)^2+\tfrac{3}{4} < \left(\frac{\xi-n}{2}\right)^2 < \left(\tfrac{1}{2}(\xi+\eta)+\tfrac{1}{2}\right)^2+\tfrac{3}{4},$$

(23b) $$(\xi-\tfrac{1}{2})(\eta-\tfrac{1}{2})+\tfrac{3}{4} < 0 < (\xi+\tfrac{1}{2})(\eta+\tfrac{1}{2})+\tfrac{3}{4}.$$

We shall first prove the right hand inequality. It is trivial if $a_{t-1} = 2$ for then $(\eta+\frac{1}{2}) > 0$. If $a_{t-1} = 1$, $-\eta = \{0, 1, \ldots\} \leqslant \{0, 1, 2, 1, 2, \ldots\} = 3^{1/2}-1$ and $\xi < \{2, 1, 2, 1, \ldots\} = 3^{1/2}+1$. Thus $(\xi+\frac{1}{2})(-\eta-\frac{1}{2}) < 9/4 - 3^{1/2} < 3/4$. To prove the left hand inequality, we just examine the less favorable case $a_t = 1$, so $\xi > \{1, 2, 1, 2, \ldots\} = (1+3^{1/2})/2$ and $-\eta > \{0, 2, 1, 2, 1, \ldots\} = (3^{1/2}-1)/2$. Thus $(\xi-\frac{1}{2})(-\eta+\frac{1}{2}) > \frac{3}{4}$. (The last inequality may appear uncomfortably "close", but in a Markoff form the ones and twos are paired, which provides stricter inequalities and hence an absolute lower bound on the distance from a point of crossing of the geodesic to the elliptic fixed points.) The exit relation is quite similar since $\xi - a_t = \{0, a_{t+1}, \ldots\}$ and $\eta - a_t = -\{a_t, a_{t-1}, \ldots\}$ so the same discussion holds, completing the proof.

7. Representation of forms by geodesics in the elliptic period plane. We now consider the geodesics transferred to the torus (or doubly periodic U-plane) by

(24) $$1 - J(z) = \wp'(U)^2 = 4\wp^3(U)+1$$

(see [3]). The period parallelogram corresponds to the fundamental domain of Γ_6 a subgroup of the modular group Γ_1 of index 6. (In [3], we saw how the Markoff triples correspond to various vector bases of the period or generating substitutions of Γ_6.) Actually the fundamental domain of the periodic structure is best represented by six image of $\boldsymbol{D}_0$, as a period hexagon (see Figure 3).

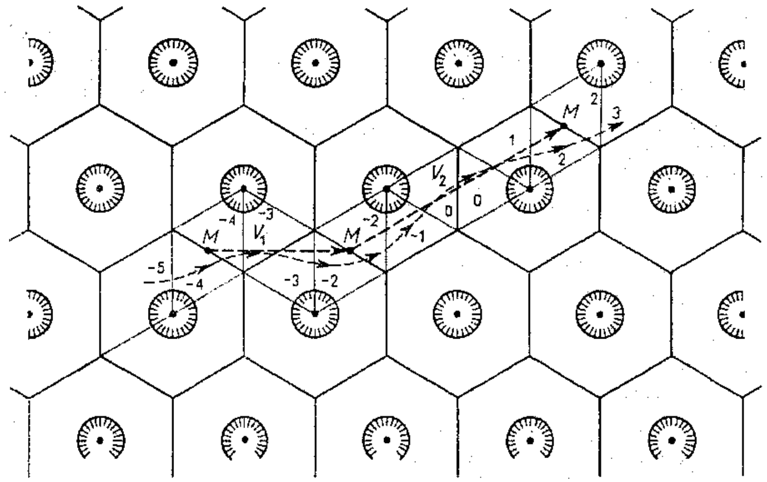

Fig. 3. Path of geodesic in U-plane, as approximated by $\{\ldots, V_1, V_2, \ldots\}$. The numbering of triangles agrees with replicas of $\boldsymbol{D}_\infty$ in Figure 2

Now the geodesic of a Markoff form will become a geodesic in the U-plane (closed under translations of the lattice periods). It will be characterized by the fact that it does not get within a certain "distance" of the images of $z = \infty$, i.e., it remains on a perforated torus with open disks excised according to the images under (24) of $\operatorname{Im} z > 3/2$.

Indeed the transformations

$$(25a)\qquad z' = 1+1/(1+1/z), \quad \text{or} \quad z' = 1+1/(\overline{1+1/\bar{z}}),$$

$$(25b)\qquad z'' = 2+1/(2+1/z), \quad \text{or} \quad z'' = 2+2/(\overline{2+1/\bar{z}})$$

correspond to displacement vectors $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$ shown in Figure 3. In transferring the geodesic from the z-plane to the U-plane it is important to note that $z+1$ represents a positive rotation within each hexagon and $1/\bar{z}$ represents a reflection across the side of the hexagon reversing the sense of rotation. Particularly the period $\{\ldots, 1, 1, \ldots, 2, 2, \ldots\}$ is homotopic to a polygon of displacement vectors $\{\ldots + \boldsymbol{V}_1 + \ldots + \boldsymbol{V}_2 + \ldots\}$ within the torus, as shown in Figure 3. (Note that this polygon completely determines the triangulation of the geodesic.)

These geodesics also have no double points on the U-plane (although they clearly would have such points on the torus); they progress monotonically in the directions $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$.

Hence, if we prescribe a (primitive) homology class on the (unperforated) torus, $v\boldsymbol{V}_1 + u\boldsymbol{V}_2$, $(u, v) = 1$ $(u \geqslant 0, v \geqslant 0)$, *there is precisely one displacement polygon (within symmetries) corresponding to that homology class and homotopic to the geodesic for a Markoff form on the perforated torus. If these displacement polygons correspond to normalized periods, then the juxtaposition of displacement polygons corresponds to the juxtaposition of periods* (in Section 4). *The closed geodesics on the perforated torus correspond to (only) the Markoff forms.*

8. Limiting geodesics. We now consider indefinite binary quadratic forms for which the coefficients are no longer proportional to integers but for which the roots are still irrational. The previous theory still holds with minor modifications. No new minima are introduced above the limit point $1/3$, but now there exist an uncountable infinity of so-called "limiting" Markoff forms where $m(Q) = 1/3$. Some partial results on these forms are cited by Koksma ([9], p. 32) and a complete discussion of the continued fraction theory appears in Dickson's account [6]. The continued fraction (7) for $\{-\eta, \xi\}$ of course is no longer periodic, but the Markoff symmetry property still holds.

What about the geodesics of the limiting Markoff forms? It follows from classical theorems of compactness of geodesics on a compact manifold that some of these geodesics can be found as limiting geodesics of integral

Markoff forms. For example, if we consider normalized periods $\boldsymbol{P}_-(u, v)$ $(u > v$, of type 2, say), we can build up the continued fraction γ of (14a) by the rules (14b) and (14d) so as to represent essentially an "irrational slope" u/v. The geodesic will have a limiting geodesic lying within the same triangulation (by the compactness theorems for compact manifolds) in the homotopy class determined by the displacement polygons (now infinite). It will not be possible in general to associate such geodesics with a "slope" u/v because the limiting ratio (of pairs 2, 2 to 1, 1) depends on the central point in the continued fraction from which we begin the averaging process.

The simplest case might be cited here, where $\gamma = (1+5^{1/2})/2$. Then we can build up the normalized periods corresponding to the convergents $\boldsymbol{P}_-(u_{n+1}, u_n)$ with $u_n = u_{n-1} + u_{n-2}$ (the Fibonacci sequence, $u_0 = u_1 = 1$). Thus we obtain

$$\begin{aligned}
(n=0)\quad & \boldsymbol{P}_-(1,1) = \{1,1,2,2\},\\
(n=1)\quad & \boldsymbol{P}_-(2,1) = \{2,2,1,1,2,2\},\\
(n=2)\quad & \boldsymbol{P}_-(3,2) = \boldsymbol{P}_-(2,1)+\boldsymbol{P}_-(1,1)\\
& \qquad\qquad = \{2,2,1,1,2,2,1,1,2,2\},\\
(n=3)\quad & \boldsymbol{P}_-(5,3) = \boldsymbol{P}_-(2,1)+\boldsymbol{P}_-(3,2)\\
& \qquad\qquad = \{2,2,1,1,2,2,2,2,1,1,2,2,1,1,2,2\}.
\end{aligned}$$

Now if we call $\boldsymbol{P}_-(u_{n+1}, u_n) = \boldsymbol{P}_n$ it is not hard to find various ways in which the periods of the continued fractions become stationary. For example, with n odd,

$$\begin{aligned}
\boldsymbol{P}_{n+2} &= \boldsymbol{P}_n + \boldsymbol{P}_{n+1},\\
\boldsymbol{P}_{n+3} &= \boldsymbol{P}_{n+2} + \boldsymbol{P}_{n+1} = (\boldsymbol{P}_n + \boldsymbol{P}_{n+1}) + \boldsymbol{P}_{n+1},\\
\boldsymbol{P}_{n+4} &= \boldsymbol{P}_{n+2} + \boldsymbol{P}_{n+3} = (\boldsymbol{P}_n + \boldsymbol{P}_{n+1}) + (\boldsymbol{P}_n + \boldsymbol{P}_{n+1} + \boldsymbol{P}_{n+1}), \text{ etc.}
\end{aligned}$$

We notice that the juxtaposition $(\boldsymbol{P}_n + \ldots + \boldsymbol{P}_{n+1}) + (\boldsymbol{P}_n + \ldots + \boldsymbol{P}_{n+1})$ is permanent. Thus the period $\boldsymbol{P}_n$ will stabilize to the right and left of each point of juxtaposition; the geodesics will converge to a limit; and the ratio of twos to ones will converge to $(1+5^{1/2})/2$. This, however, is a very special situation.

It would be interesting to know if all of the limiting Markoff forms (with $m(Q) = 1/3$) can be determined from some type of continued fraction device. Very little has been done with such problems, and insufficient use has been made of geometric representations (compare [1] for a related ergodic problem).

References

[1] E. Artin, *Ein mechanische System mit quasiergodischen Bahnen*, Abh. Math. Sem. Hamburg (1924), pp. 170–175.
[2] J. W. S. Cassels, *The Markoff chain*, Ann. of Math. 50 (1949), pp. 676–685.
[3] H. Cohn, *Approach to Markoff's minimal forms through Modular Functions*, Ann. of Math. 61 (1955), pp. 1–12.
[4] H. Davenport, *Note on the product of three homogeneous linear forms*, Journ. London Math. Soc. 16 (1941), pp. 98–101.
[5] — *On the product of three homogeneous linear forms IV*, Proc. Cambridge Philos. Soc. 39 (1943), pp. 1–21.
[6] L. E. Dickson, *Studies in the theory of numbers*, Chicago (1930), pp. 79–107.
[7] G. Frobenius, *Über die Markoffschen Zahlen*, Preussische Akad. Wiss. Sitzungsberichte, (1913), pp. 458–487.
[8] F. Klein, *Vorlesungen über die theorie der elliptischen Modulfunktionen, I*, Leipzig (1890), Vol. 1, p. 220.
[9] J. F. Koksma, *Diophantische Approximationen*, Berlin 1936, p. 33.
[10] A. A. Markoff, *Sur les formes binaires indéfinies*, Math. Ann. 15 (1879), pp. 381–409.
[11] L. J. Mordell, *The product of three homogeneous linear ternary forms*, Journ. London Math. Soc. 17 (1942), pp. 107–115.
[12] C. J. Hightower, *Minima of indefinite quadratic forms*, Journ. Number Theory 2 (1970), pp. 364–378.

MATHEMATICS DEPARTMENT
UNIVERSITY OF ARIZONA
Tucson, Arizona

Received on 24. 1. 1970

ACTA ARITHMETICA
XVIII (1971)

On the least integers represented by the genera of binary quadratic forms

by

A. Baker (Cambridge) and A. Schinzel (Warszawa)

In memory of our teacher and friend
Harold Davenport

1. Introduction. It is well known that every class of primitive binary quadratic forms with discriminant D represents a positive integer $\leqslant c\sqrt{|D|}$, where c is an absolute constant. In general the estimate cannot be improved either for $D<0$ or for $D\geqslant 0$; in the first case this is shown by the sequence $nx^2+(n+1)y^2$ $(n=1,2,\ldots)$ and in the second it follows as a consequence of the so-called Markov chain. The main aim of this paper is to show that for genera of forms in question, however, a better estimate obtains:

Theorem 1. *Every genus of primitive binary quadratic forms with discriminant D represents a positive integer $\leqslant c(\varepsilon)|D|^{3/8+\varepsilon}$ for any $\varepsilon>0$, where $c(\varepsilon)$ denotes a number depending only on ε.*

It seems likely that in fact every genus represents a positive integer $\leqslant c(\varepsilon)|D|^{\varepsilon}$ for any $\varepsilon>0$ but we are unable to establish this conjecture with our present arguments. Also our proof does not allow the number $c(\varepsilon)$ mentioned in the theorem to be effectively computed when $\varepsilon<1/8$; an effective estimate, on the other hand, would enable one to determine all the "numeri idonei" of Euler. This results at once from the following:

Theorem 2. *All negative discriminants D with one class of forms in every genus satisfy $|D|\leqslant C(\varepsilon)$, where $C(\varepsilon)$ is effectively computable in terms of $c(\varepsilon)$ for any $\varepsilon<1/8$.*

The proof of Theorem 1 involves an argument similar to that used by Linnik and Vinogradov in their paper [8] on the least prime quadratic residue; thus we shall appeal to the well-known result of Burgess [3] on character sums and also to Siegel's fundamental theorem [12] on L-functions. The proof of Theorem 2 is based on the work of [2]. As demonstrated there, for negative fundamental discriminants D with class number 1 or 2, where, in the latter case, D is assumed to be even, the integers referred to in Theorem 1 can be given explicitly and so, by