Markoff's Numbers and Quadratic Forms

G.V.Belyi

April 25, 1997

Introduction

The natural solutions of the Markoff equation $x^2 + y^2 + z^2 = 3xyz \quad (x \geq y \geq z)$ are called the *"Markoff numbers"*. A.A.Markoff investigated this equation in [1] , while studying rational approximations of real numbers. There is a conjecture asserting that any solution of the Markoff equation is determined by x. This assertion is often refered to as the uniqueness of the Markoff numbers. Recently the interest to this topic has grown in connection with the A.L.Gorodentsev and A.N.Rudakov works [6], [7], where they had proved that the dimensions of exceptional bundles on $\mathbb{P}^2$ are Markoff's numbers.

The aim of this article is to explain some Markoff's numbers properties in a very explicit way. We give another proof of some results of Hurwitz and Frobenius [2], [3] relating to the Markoff numbers. It seems that the construction in our proof can be modified to study this equation over the ring of Gaussian integers for example. Next, we obtain a certain convex curve like a level line and discuss the Markoff numbers growth. Finally, we add some speculations relating to the uniqueness conjecture.

I am grateful to S. Kuleshov, B.Karpov, and D.Orlov for valuable discussions and the invaluable help with LaTeX . I am also grateful to A.N.Tyurin for his stimulating interest in my work. I would like to thank the MPI for hospitality.

1 Preliminary

1.1 For any natural number $S \geq 3$, we choose the unit

$$\varepsilon = \frac{S + \sqrt{(S-2)(S+2)}}{2} = \frac{S + k\sqrt{D_0}}{2},$$

where D_0 is a square-free integer. The index f_0 of the order $\mathfrak{O}_1 = \mathbb{Z}[\varepsilon]$ in the maximal order $\mathfrak{O}$ of the field $\mathbb{Q}(\sqrt{D_0})$ is equal to k whenever $D_0 \equiv 1\,(mod\,4)$. In other cases $(D_0 \equiv 2, 3\,(mod\,4))$ these index f_0 is equal to $k/2$. For each divisor f of the index $f_0 = f f_1$ there exist an unique intermediate order $\mathfrak{O}_f$ such that

$$[\mathfrak{O}_f : \mathfrak{O}_1] = f, \qquad [\mathfrak{O} : \mathfrak{O}_f] = f_1.$$

1.2 For all these orders, we consider the corresponding class groups $\widehat{Cl}$ of oriented modules M up to strict similarity; this means that the modules M and αM are equivalent if the norm of $\alpha \in \mathbb{Q}(\sqrt{D_0})$ is positive. The orientation is a choice of such a basis $\langle \lambda, \mu \rangle = M$ of the module that the determinant

$$\det \begin{pmatrix} \lambda & \mu \\ \tilde{\lambda} & \tilde{\mu} \end{pmatrix} > 0 \quad ,$$

were $\sim$ denotes the conjugation. Moreover, we can choose the second basis element with positive norm and divide by it. So, without loss of generality, we suppose it is 1. We refer to the textbook [4] for further details. Let us choose the basis for the main modules $\mathfrak{O}_f$ as follows: $\mu = 1$, $\lambda = (\varepsilon - d)/f$, where d is a rational integer solution of the congruense $2d \equiv S\,(mod\; f)$. Then the matrix representing the unit ε on the main module $\mathfrak{O}_f$ has the following form

$$A_1 = \begin{pmatrix} fa+d & fbc \\ f & d \end{pmatrix},$$

and $\lambda^2 = a\lambda + bc$. When we consider the action of the unit ε on an arbitrary module M with the coefficient ring $\mathfrak{O}_f$ it's matrix will be of the form

$$A = \begin{pmatrix} fa+d & fb \\ fc & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$$

with $2d + fa = S$ and $\gcd(a,b,c) = 1$. The corresponding module $M = \langle \mu, 1 \rangle = \langle \lambda/c, 1 \rangle$ is determined by the matrix A : μ is a root of the quadratic equation $c\mu^2 - a\mu - b = 0$, or in other words, it is a fixed point of the linear-fractional transform which corresponds to the matrix A. Indeed,

$$\varepsilon \cdot 1 = \frac{fa + 2d + \sqrt{f^2a^2 + (4fad + 4d^2 - 4)}}{2} = fc\frac{a + \sqrt{a^2 + 4bc}}{2c} + d = fc \cdot \mu + d \cdot 1.$$

Thus we obtain the correspondence between classes of modules and the conjugation classes of matrices in the group $\mathrm{SL}(2,\mathbb{Z})$.

1.3 We need to find a matrix $\tilde{A}$ corresponding to the module $\tilde{M}$ inverse to the M. It is proved in [4] that the module $\tilde{M} = \langle -\tilde{\mu}, 1 \rangle$ is such a module. Obviously, it it is positively oriented, so

$$\tilde{A} = \begin{pmatrix} d & fb \\ fc & fa+d \end{pmatrix}$$

Next we find the composition rule for modules over the same order:

1.4 Lemma. *Suppose that the matrices*

$$\begin{pmatrix} fa+d & fbc_2 \\ fc_1 & d \end{pmatrix}, \quad \begin{pmatrix} fa+d & fbc_1 \\ fc_2 & d \end{pmatrix}, \quad \begin{pmatrix} fa+d & fb \\ fc_1c_2 & d \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$$

correspond to three modules M_1, M_2, and M_3 over the same order, then the third module is the composition of the first and the second one.

Proof. We have $M_1 = \langle \lambda/c_1, 1 \rangle, M_2 = \langle \lambda/c_2, 1 \rangle, M_3 = \langle \lambda/c_1c_2, 1 \rangle$.

$$M_1 * M_2 = \langle 1, c_1\frac{\lambda}{c_1c_2}, c_2\frac{\lambda}{c_1c_2}, \frac{\lambda^2}{c_1c_2} \rangle = \langle 1, c_1\frac{\lambda}{c_1c_2}, c_2\frac{\lambda}{c_1c_2}, a\frac{\lambda}{c_1c_2} + b \rangle = M_3,$$

as $\gcd(a, c_1, c_2) = 1$.

1.5 Definition. *A matrix* $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}(2,\mathbb{Z})$ *with* $tr\,A \geq det\,A + 2$ *is called reduced if* $\quad a \geq b,\ c \geq d \geq 0.$

1.6 Lemma.

i) Each conjugation class of matrices with trace $tr\,A \geq det\,A + 2$ *contains a reduced matrix* $\bar{A}$.

ii) For the module $M = \langle \mu, 1 \rangle$ *associated to the reduced matrix* $\bar{A}$, $\mu > 1$ *and* $-1 < \tilde{\mu} < 0$.

Proof. If the absolute value $|d| > |b|$, let $\bar{d}$ be a residue of d modulo $|b|$. The matrices $A \sim \begin{pmatrix} a+d-\bar{d} & b \\ * & \bar{d} \end{pmatrix}$ are conjugated in $\mathrm{SL}(2,\mathbb{Z})$. The rest is evident.

1.7 The Euclidean algorithm provides for each reduced matrix $A \in \mathrm{SL}(2,\mathbb{Z})$ the unique decomposition into the product :

$$A = \begin{pmatrix} a_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} a_2 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} a_{2m} & 1 \\ 1 & 0 \end{pmatrix} \quad , \quad a_i \in \mathbb{N},$$

and this decomposition is nothing else as the continued fraction for the rational number a/c, or some period of the continued fraction for the irrational number μ. We write, for brevity, $A = [a_1, a_2, \ldots, a_{2m}]$. Henceforth we do not distinguish the matrices and the associated modules. Consider the cyclic permutation $\sigma = (1, 2, \ldots, 2m)$. It acts on the period A in an evident way. And each even power of this action gives us a module equivalent to A. Moreover, each equivalent reduced module is obtained in such a way. Note that 1.3 gives us that the inverse module $\tilde{A} \sim [a_1, a_{2m}, \ldots, a_3, a_2]$.

1.8 Consider now the natural epimorphism $\psi : \widehat{Cl} \to Cl$ onto the usual ideal class group of the order $\mathfrak{O}_f$ which forgets the orientation and changes the strict similarity to usual one. If the order $\mathfrak{O}_f$ contains an unit with the norm -1 , then the epimorphism ψ is an isomorphism. Otherwise, the kernel of this epimorphism ψ is the group $\{\pm 1\}$. Denote by A_{-1} the generator of this kernel. So $\ker(\psi) = \{A_1, A_{-1}\}$. If we take the module A as in 1.7 , then $A * A_{-1} \sim [a_2, a_3, \ldots, a_{2m}, a_1]$. It follows easily from 1.4 .

1.9 Definition. *The module* A *is said to be ambiguous if* $A * A \sim A_1$, *i.e., if it is of the second order. We shall call the module* A *semi-ambiguous if* $A * A \sim A_{-1}$ *or equivalently* $\tilde{A} \sim A * A_{-1}$,

1.10 Lemma. *Each semi-ambiguous module is equivalent to the symmetric module* $[a_1, a_2, \ldots, a_m, a_m, \ldots, a_2, a_1]$ *and each ambiguous module is equivalent to the symmetric module* $[a_1, a_2, \ldots, a_m, a_{m+1}, a_m \ldots, a_2]$.

Proof. Making use of 1.7 and 1.8 we obtain the equality

$$[a_{2m}, \ldots a_2, a_1]^{\tau} \;=\; [a_1, a_2, \ldots, a_{2m}],$$

where τ is an even power of σ for semi-ambiguous modules and is an odd power of σ for ambiguous modules. Since a product of a reflection and a rotation is a reflection again, this concludes the proof.

Note that if we started with the unit $\epsilon = \frac{T+\sqrt{T^2+4}}{2}$ with norm -1, then the notions of ambiguous modules and semi-ambiguous modules coincide. In this case, each ambiguous module is equivalent to the symmetric module $[a_1, \ldots, a_m, \ldots, a_1]$.

1.11 Note that for matrices $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $B = \begin{pmatrix} m & * \\ n & * \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$ we have

$$B^{-1} \cdot A \cdot B = \begin{pmatrix} * & * \\ t & * \end{pmatrix}, \text{ where } t = cm^2 - (a-d)mn - bn^2.$$

2 Markoff's numbers and semi-ambiguous forms

2.1 Now, let (x, y, z, p) be a natural solution of the following equation

$$x^2 + y^2 + z^2 = pxyz$$

and suppose y, z are coprime. Than there exist a matrix $B = \begin{pmatrix} y & * \\ z & * \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$ such that
$B^{-1} \cdot \begin{pmatrix} px & -1 \\ 1 & 0 \end{pmatrix} \cdot B = \begin{pmatrix} a & b \\ -x^2 & d \end{pmatrix}$, see 1.11 . Note that the matrix $\begin{pmatrix} px & -1 \\ 1 & 0 \end{pmatrix}$ is equivalent to the matrix A_1 from 1.2 and corresponds to the main module $\mathfrak{O}_1 = \mathbb{Z}[\varepsilon]$ with trace $S = px = a + d$.

2.2 In the cases $x = 1$ or $x = 2$ the equation 2.1 turns in the *"negative Pell's equation"* and means that the order $\mathfrak{O}_2$ contains an unit with negative norm. Denote this unit by

$$\epsilon = \frac{T + \sqrt{T^2 + 4}}{2}. \qquad \text{We have} \qquad \epsilon^2 = \frac{T^2 + 2 + T\sqrt{T^2 + 4}}{2}.$$

As this natural number T divides $f = 2$, we obtain $T = 1,\ f = 1,\ S = 3,\ p = 3$, for $x = 1$, and $T = 2$, $f = 2,\ S = 6,\ p = 3$ for $x = 2$. Henceforth we assume that $x > 2$.

2.3 Lemma. *The matrix* $\begin{pmatrix} a & -xb \\ x & d \end{pmatrix} \in \mathrm{SL}(2, \mathbb{Z})$ *is semi-ambiguous and therefore it is conjugated with a reduced symmetrical matrix* $A = L^T \cdot L$ *by Lemma 1.10 .*

Proof. Let $f = \gcd(x, a - d)$ be the corresponding index. If $f = 1$, then the lemma follows from 2.1 and 1.4 .

Note that the index f can be at most 2, as x divides $S = a + d$. In this case the numbers y, z, a, d have to be odd, $ad \equiv 1\,(mod\,4)$, $S \equiv 2\,(mod\,4)$, and we can write $x = 2x_0$, with odd x_0. Now the matrix $A_2 = \begin{pmatrix} S-1 & (S-2)/2 \\ 2 & 1 \end{pmatrix}$ corresponds to the main module $\mathfrak{O}_2$. We set $C = \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \cdot B \cdot \begin{pmatrix} 1/2 & 0 \\ 0 & 1 \end{pmatrix}$. It is easy to check that $C \in \mathrm{SL}(2, \mathbb{Z})$ and $C^{-1} \cdot A_2 \cdot C = \begin{pmatrix} a & 2b \\ -2x_0^2 & d \end{pmatrix} \sim A * A * A_{-1}$. We denote by α the symmetry axis.

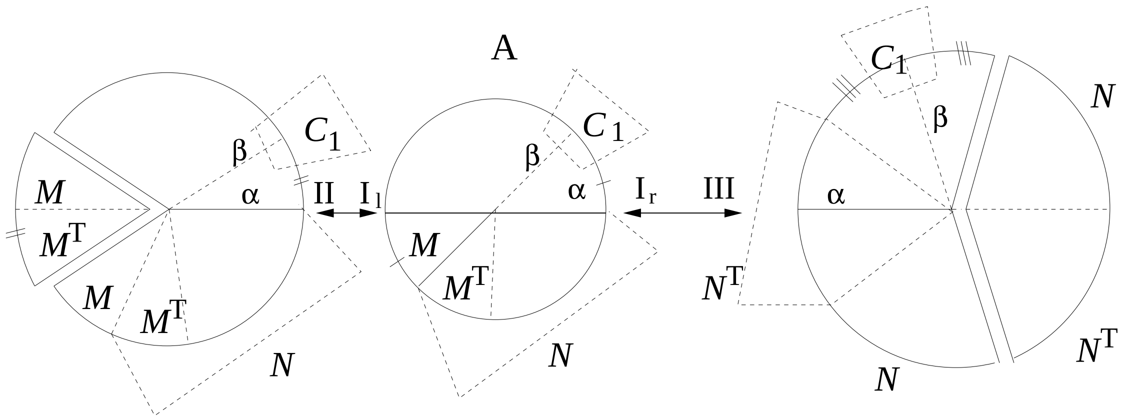
2.4 Next we show that matrix A from 2.3 has another axis β of the "almost" symmetry. Let $\bar{d}$ be a residue of d modulo x . We can assume without loss of generality that $0 < \bar{d} < x/2$, otherwise we change A to $\tilde{A}$, see 1.3 . Further, consider the following product

$$C=\begin{pmatrix} p & -1\\ 1 & 0\end{pmatrix}^{-1}\cdot\begin{pmatrix} px-\bar d & *\\ x & \bar d\end{pmatrix}=\begin{pmatrix} x & \bar d\\ \bar d & *\end{pmatrix}\in \mathrm{SL}(2,\mathbb{Z}).$$

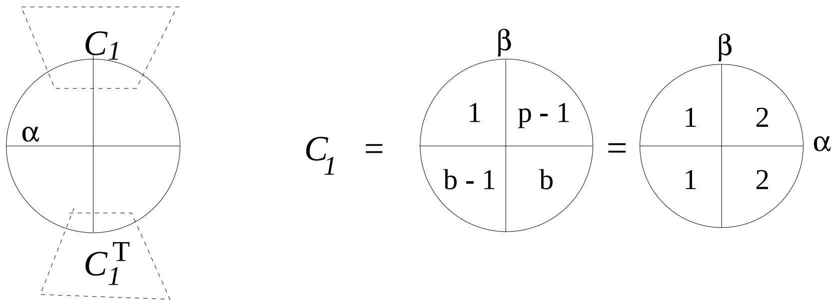
We have the reduced symmetrical matrix C. Taking into account 1.7 , we obtain $C=[b,b_1,\dots,b_n,b_n,\dots,b_1,b]$. The condition $\bar d<x/2$ gives us $b\geq 2$. We see that the matrix $\tilde A$ is equivalent to the usual product $C_1\cdot C_2$, where C_2 is equal to either the identity matrix Id whenever $n=0$ or to $C_2=[b_1,\dots,b_n,b_n,\dots,b_1]$, and $C_1=\begin{pmatrix} b & 1\\ 1 & 0\end{pmatrix}\cdot\begin{pmatrix} p & -1\\ 1 & 0\end{pmatrix}\cdot\begin{pmatrix} b & 1\\ 1 & 0\end{pmatrix}$. So we have $C_1=[b+1,b-1]$ for $p=1$ and $C_1=[b,p-1,1,b-1]$ for $p>1$. This other axis β gives us the following.

2.5 Theorem. *(Markoff, Hurwitz) The equation from 2.1 has solutions with coprime y,z iff $p=3$. In these case the equation turns in* "Markoff's equation" *and we have $3x=trA$ for the matrix A from 2.4 .*

PROOF. Let the matrix A have two such axes. We arrange its factors (see 1.7) equidistantly along a circle. Consider two transformations II , III of the matrix A , as it is represented on the pictures



, where C_1, α, β are as in 2.4 . The other notations are evident. It is clear that these transformations give the new matrixes with the two such axes and have the inverse; denote them by I_l , I_r . We can write $L_x=M\cdot N$ and denote by m , n the quantities of factors in M , N . If we only use the transformations I_l , I_r the Euclidean algorithm provides us, the axes α and β become orthogonal after finitely many steps. We see on the next picture



that the upper semicircle can not contain the matrix C_1 as C_1^T is not symmetrical. This means that $b=2$, $p=3$, and m , n are coprime; and we obtain Markoff's number 5 corresponding to the matrix $C_1=[2,2,1,1]$.

2.6 For a solution (x, y, z) with $x > y > z$ the transformation II corresponds to the mutation $(x, y, z) \to (3xz - y, x, z)$, and the transformation III corresponds to the mutation $(x, y, z) \to (3xy - z, x, y)$. The transformations I corresponds to the mutation $(x, y, z) \to (y, z, 3yz - x)$; in more detail, we have

$$L_z = M^T\,,\ L_y = M\,N\,,\ L_x = M^T\,M\,N\,,\ L_{\tilde{x}} = N \quad \text{for } \mathrm{I_l} \text{ and}$$

$$L_y = M\,N\,,\ L_z = N^T\,,\ L_x = M\,N\,N^T\,,\ \ L_{\tilde{x}} = M \quad \text{for } \mathrm{I_r}\,.$$

2.7 Having applied the transformations I to the matrix C_1, we obtain the main modules $[1, 1]$ and $[2, 2]$ corresponding to the singular Markoff numbers 1 and 2 . We denote them

$$X = [1, 1] = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \quad \text{and} \quad Y = [2, 2] = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}.$$

It is not clear how we could define the “almost” symmetry axis β for them. So the matrix C_1 can be restored explicitly neither by X, nor by Y.

2.8 Next we find the relations between quantities of the twos, denote them by k, k_m, k_n , and the ones, denote them by l, l_m, l_n , in the decompositions of L, M, N. Denote by m, n the quantities of factors in $M,\ N$. It is clear that $k + l = m + n$. Show that

$$\det \begin{pmatrix} l_m & l_n \\ k_m & k_n \end{pmatrix} = 1 \quad ,$$

For $L = [1, 2]$ we have $\det Id = 1$. The transformation II corresponds to addition of the first column to the second and the transformation III corresponds to addition of the second column to the first one. We obtain

$$km - 1 = (k_m + k_n)(k_m + l_m) - l_m k_n + l_n k_m = k_m(k_m + k_n + l_n + l_m) = k_m(m + n)$$

$$\text{or} \qquad km \ \equiv \ ln \ \equiv \ 1 \ \ (mod \ (m + n))\,.$$

Now we can consider the pair (k, l) as the coordinates for the Markoff number x.

2.9 Examples:

$tr[1, 1] = 3 \cdot 1(0, 1)$, $tr[2, 2] = 3 \cdot 2(1, 0)$, $tr[2, 1, 1, 2] = 3 \cdot 5(1, 1)$,
$tr[2, 1, 1, 1, 1, 2] = 3 \cdot 13(1, 2)$, $tr[1, 2, 2, 2, 2, 1] = 3 \cdot 29(2, 1)$,
$tr[2, 1, 1, 1, 1, 1, 1, 2] = 3 \cdot 34(1, 3)$, $tr[2, 1, 1, 1, 1, 1, 1, 1, 1, 2] = 3 \cdot 89(1, 4)$,
$tr[1, 2, 2, 2, 2, 2, 2, 1] = 3 \cdot 169(3, 1)$, $tr[1, 2, 2, 1, 1, 1, 1, 2, 2, 1] = 3 \cdot 194(2, 3)$,
$tr[2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2] = 3 \cdot 233(1, 5)$, $tr[2, 1, 1, 2, 2, 2, 2, 1, 1, 2] = 3 \cdot 433(3, 2)$.

Compare [3].

2.10 It is obvious that either A or $\tilde{A}$ is equivalent to the product of k pieces of Y and l pieces of X in spite of our axes can cut some factors. Let us imagine each factor to be composed of two equal pieces like an oxygen molecule O_2. Taking into account that after each transformation the separated “oxygen atoms” unite again into “oxygen molecules”, we see this.

The proof of theorem 2.5 gives an answer to the following question: there are kr pieces of the matrix Y and lr pieces of the matrix X with coprime $k,\ l$; how to intermix them in maximal uniform way to the product having minimal trace? The answer is: A^r is such a product.

Besides, we obtain a possibility to compute the Markoff number $x(k, l)$ without computing the intermediate numbers. We divide a circle into $2(k+l)$ equal pieces. Next we solve the congruenses

$$km \equiv ln \equiv 1 \ (mod\ (k+l))$$

by using the Euclidean algorithm. Of course, we obtain the coordinates for the intermediate numbers, but not the numbers itself. Further, we draw the axes α, β, and arrange the twos and the ones along the circle. Example: Let $k=7$, $l=10$; then $m=5$, $n=12$, and $A = YXYXYXXYXYXXYXYXX$, $3x(k, l) = trA$.

2.11 Let the matrices M, N be as in 2.5 . The "almost" symmetry axis β provides us $NN^TM^TM = K^TC_1K$ for some $K \in \mathrm{GL}(2,\mathbb{Z})$. We have

$$C_1 = [2,2,1,1] = \begin{pmatrix} 12 & 7 \\ 5 & 3 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} + 3\begin{pmatrix} 2 & 1 \end{pmatrix}^T \begin{pmatrix} 2 & 1 \end{pmatrix}.$$

Write $(r\ s) = (2\ 1)K$. As all the elements of the group $\mathrm{GL}(2,\mathbb{Z})$ are conformally symplectic, we obtain

$$K^TC_1K = \begin{pmatrix} 3r^2 & 3rs \pm 1 \\ 3rs \mp 1 & 3s^2 \end{pmatrix},$$

and $x = r^2+s^2$. Besides, $NN^TM^TM = \pm W + S$, where $W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ and S is a symmetrical matrix, compare [3]. It is easily shown that $\gcd(x, y+zi) = r \pm si$, and

$$Y \cdot K^T \cdot K \cdot X = \begin{pmatrix} 3x - \bar{d} & * \\ x & \bar{d} \end{pmatrix}.$$

3 Polynomial solutions of a pseudo-Markoff equation

3.1 The straight lines on the Markoff cubic are defined over the integer Gaussian numbers ring $\mathbb{Z}[i]$ except three lines at infinity. To stay over $\mathbb{Z}$ we change the equation 2.1 to

$$x^2 - y^2 - z^2 = \pm pxyz.$$

For a natural solution (x,y,z,p) with coprime x,y,z we have

$$\begin{pmatrix} a & b \\ x^2 & d \end{pmatrix} \sim \begin{pmatrix} px & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} a' & b' \\ y^2 & d' \end{pmatrix} \sim \begin{pmatrix} py & 1 \\ 1 & 0 \end{pmatrix},$$

see 2.1 . Repeating the arguments in 2.3 and 2.4 , we obtain a symmetry axis α and a "almost" symmetry axis β for the matrices

$$A \sim \begin{pmatrix} a & xb \\ x & d \end{pmatrix} \quad \text{and} \quad B \sim \begin{pmatrix} a' & yb' \\ y & d' \end{pmatrix} \quad \text{again}.$$

3.2 There is a slight difference with 2.5 : both matrixes A and B are ambiguous, so these axes can cut some factors. We can repeat the arguments in theorem 2.5 taking into account 2.10 . Using the first transformation we obtain 3 simplest forms for the matrix A with ortogonal axes α, β

i) $A = [1, b-1, 1, 1, b, 1]$, $p = 2$,

ii) $A = [b-1, b+1]$, $p = 1$,

iii) $A = [1, p-1]$.

In case i) we can take

$$2x = trA = 2(2b^2 + 2b + 1), \quad 2y = tr[1, b, 1] = 2(b+1), \quad 2z = tr[1, b-1, 1] = 2b$$

We see that this case reduces to the second one, where

$$x = trA = b^2 + 1, \qquad y = tr[b+1] = b+1, \qquad z = tr[b-1] = b-1,$$

by a linear change of variables. Both these cases correspond to the straight line

$$y = t+1, \;\; z = t-1, \;\; x = 2 \quad \text{on the cubic} \quad y^2 + z^2 - x^2 = xyz\,.$$

In case iii) we have singular solution like in 2.7 .

The matrix $[1, p-1]$ can be obtained by the transformation I from the two different matrixes $[p, p-1, 1, p-1]$ or $[1, p-1, 1, 0] \sim [2, p-1]$ with the axes angle 45^o. As above, the second case reduces to the first one by a linear change of variables. In the first case we have

$$px = tr[p, p-1, 1, p-1] = p(p^2+1), \;\; py = tr[p-1, 1, p-1] = p^2, \;\; pz = tr[p] = p\,.$$

We obtain the solution of 3.1 corresponding to the straight line

$$x = t, \;\; y = 0, \;\; z = t \quad \text{on the same cubic} \quad y^2 + z^2 - x^2 = xyz\,.$$

It is not hard to check that the other straight lines on the cubic are symmetrical to these two or lie at infinity. Compare [8].

4 Markoff's numbers growth

4.1 Let (x, y, z) be a real solution of the equation

$$x^2 + y^2 + z^2 = xyz\,,$$

where $x \geq y \geq z > 2$. There exist positive $\alpha > \beta > \gamma$ such that

$$x = 2\cosh\alpha\,, \;\; y = 2\cosh\beta\,, z = 2\cosh\gamma\,, \text{ and}$$
$$\cosh^2\alpha + \cosh^2\beta + \cosh^2\gamma = 2\,\cosh\alpha\,\cosh\beta\,\cosh\gamma\,.$$

For the mutation $\tilde{x} = yz - x$ we denote by $\tilde{\alpha}$ the corresponding logarithm, so

$$\cosh\alpha + \cosh\tilde{\alpha} = 2\,\cosh\beta\,\cosh\gamma\,.$$

4.2 Let $\mathfrak{L}$ be a lattice with two generators $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{a} \wedge \mathbf{b} = 1$. Denote the origin of the lattice $\mathfrak{L}$ by O . We arrange the logarithms α, β, γ and all their mutations in the lattice $\mathfrak{L}$ in the following way. We put γ at the vertices $\pm\mathbf{a}$; we put β at the vertices $\pm\mathbf{b}$; and we put α at the vertices $\pm(\mathbf{a}+\mathbf{b})$. Further, suppose we have a fundamental parallelogram of the lattice $\mathfrak{L}$ with three vertices being filled, for example, $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{a}+\mathbf{b}$. Take the diagonal of the parallelogram that contains the origin O . It divides the parallelogram into two triangles. Reflect these triangles to obtain two new parallelograms with empty vertices. Put the mutations at these empty places to get triples of logarithms correspondig to solutions of 4.1 at the parallelogram's vertices. In our example we put the mutation $\tilde{\beta}$ at the vertex $2\mathbf{a}+\mathbf{b}$ and we put the mutation $\tilde{\gamma}$ at the vertex $2\mathbf{b}+\mathbf{a}$. We do the same for the remaining old parallelograms. Next we do the same for all new parallelograms and so on. Each primitive point of the lattice $\mathfrak{L}$ will be filled by some logarithm. Further, we fill nonprimitive points in the following way. We put $n\eta$ at the vertex $n\mathbf{c}$ if n is natural, $\mathbf{c}$ is primitive, and η corresponds to $\mathbf{c}$.

Now we move all the points of the lattice $\mathfrak{L}$ towards the origin O or away from the origin, except the point O itself as follows. We move $\mathbf{c}$ to $m(\mathbf{c})=\mathbf{c}/\eta$, if η corresponds to $\mathbf{c}$.

4.3 Theorem. *There exists a convex curve Γ that contains all the points $m(\mathbf{c})$.*

Proof. We obtain the curve Γ as a limit sequence of convex $(6\cdot 2^n)$ -gons. First we have to check the hexagon we start with is convex. If we had $\alpha=\beta+\gamma,$ the point $m(\mathbf{a}+\mathbf{b})$ would be located on the segment $m(\mathbf{a})\,m(\mathbf{b})$. But we have

$$\cosh\alpha=\cosh\beta\cdot\cosh\gamma+\sqrt{\sinh^2\beta\cdot\sinh^2\gamma-1}\;<\;\cosh(\beta+\gamma)\,,$$

see 4.1 . This inequality means that the point $m(\mathbf{a}+\mathbf{b})$ is situated behind the straight line $m(\mathbf{a})\,m(\mathbf{b})$ with respect to the origin O.

Next we consider the parallelogram O, $(n+1)\mathbf{a}+n\mathbf{b}$, $(2n+1)(\mathbf{a}+\mathbf{b})$, $(n+1)\mathbf{b}+n\mathbf{a}$ with area $2n+1$. We have

$$t=2\cosh\,(2n+1)\alpha=tr\ C^{2n+1}\,,\qquad\text{where}\quad C=\begin{pmatrix}x&-1\\1&0\end{pmatrix}.$$

Let $\tilde{\tilde{\beta}},\tilde{\tilde{\gamma}}$ correspond to the other vertices of the parallelogram. We have to check that $\tilde{\tilde{\beta}}+\tilde{\tilde{\gamma}}>(2n+1)\alpha.$ It can be easily shown that

$$t_1=2\cosh\tilde{\tilde{\beta}}=tr\ C^n\cdot\begin{pmatrix}y&0\\z&0\end{pmatrix}\qquad\text{and}\qquad t_2=2\cosh\tilde{\tilde{\gamma}}=tr\ C^n\cdot\begin{pmatrix}z&0\\y&0\end{pmatrix}.$$

So we have to obtain the inequality

$$4+t\,t_1\,t_2>t^2+t_1^2+t_2^2\,,$$

which gives us

$$\frac{t_1}{2}\frac{t_1}{2}+\sqrt{\frac{t_1^2}{4}-1}\sqrt{\frac{t_2^2}{4}-1}>\frac{t}{2}\,.$$

It is clear that $C^n=\mu\cdot Id-\lambda\cdot C^{-1}$ with positive μ,λ. We have

$$C^{2n+1}=\mu^2C-2\mu\lambda Id+\lambda^2C^{-1},\quad t=x(\mu^2+\lambda^2)-4\mu\lambda,\quad t_1=\mu y-\lambda z,\quad t_2=\mu z-\lambda y.$$

As $y^2+z^2=x\tilde{x},\quad yz=x+\tilde{x},\quad \mu^2+\lambda^2-x\mu\lambda=1,$ we obtain

$$t_1t_2=(\mu^2+\lambda^2)yz-\mu\lambda(y^2+z^2)=(\mu^2+\lambda^2)(x+\tilde{x})-\mu\lambda x\tilde{x}=\tilde{x}+(\mu^2+\lambda^2)x\,,$$

$$t_1^2+t_2^2=(\mu^2+\lambda^2)(y^2+z^2)-4\mu\lambda yz=(\mu^2+\lambda^2)x\tilde{x}-4\mu\lambda(x+\tilde{x})=\tilde{x}t+4\mu\lambda x\,.$$

Substituting these expressions in our inequality, we see that it means that

$$4>t[t-(\mu^2+\lambda^2)x]+4\mu\lambda x \qquad \text{and} \qquad 1>-\mu\lambda(t-x).$$

4.4 Let us put the Markoff numbers in the their growth order. The theorem means that we can regard the convex curve $\alpha\Gamma$ as a level line. It bounds the domain containing the points of the lattice $\mathfrak{L}$ with the logarithms smaller than α, the points with the logarithms bigger than α lie outside, and the points with the logarithms equal to α lie on the curve. The number of the primitive points belonging to the domain bounded by the curve $\alpha\Gamma$ is proportional to area of the domain ,i.e., it is proportional to α^2, up to $o(\alpha^{1+\varepsilon})$. It means that the number of the Markoff numbers x with $3x\leq 2\cosh\alpha$ is equal to

$$C\alpha^2+o(\alpha^{1+\varepsilon}) \qquad (\alpha\to\infty) \qquad \text{with some constant } C\,, \qquad 0.18061<C<0.18076\,.$$

To obtain the interval for C we find the area of 24-gon inscribed into Γ and we find the area of 24-gon described around Γ ; then we multiply them by Chebyshev's $6/\pi^2$ and divide by 12. Note that Zagier, [5], used similar arguments and found the constant C without explicit use of this curve Γ, let alone its convexity. In the same article an estimate

$$\sim e^{c_1\sqrt{N}+c_2}\,,$$

with some constants c_1, c_2 for the Markoff number enumerated by N was given.

5 Discussion of the uniqueness conjecture

5.1 It seems to be very strange, but no correct proof of this conjecture has been advanced until now, in spite of the fact that the Markoff numbers grow so fast.

Clearly, the triple $(x>y>z)$ is restored by the couple (r,s) with $x=r^2+s^2$, see 2.11 . We have the similar situation for $3x-2$ and $3x+2$. For each element $L\in\mathrm{GL}(2,\mathbb{Z})$ we can write

$$2L=(L+L^W)+(L-L^W)=\begin{pmatrix} a & b\\ -b & a\end{pmatrix}+\begin{pmatrix} c & d\\ d & -c\end{pmatrix},$$

so $$4\,detL = a^2-c^2-d^2+b^2\,, \qquad 2\,trL\,L^T=a^2+c^2+d^2+b^2\,, \qquad \text{and}$$

$$a^2+b^2=trL\,L^T+2\,detL\,, \qquad c^2+d^2=trL\,L^T-2\,detL\,.$$

Let us return to the notations in the section 2. Taking L_x as in 2.5 , we obtain

$$a^2+b^2=3\,x\;\pm\,2\,, \qquad c^2+d^2=3\,x\;\mp\,2\,.$$

Denote by $\tilde{a}$, $\tilde{b}$, $\tilde{c}$, $\tilde{d}$ the corresponding integers for $\tilde{x}$. Combining 2.6 and 2.11 , we see

$$\det\begin{pmatrix} a & b\\ \tilde{a} & \tilde{b}\end{pmatrix}=\det\begin{pmatrix} c & d\\ \tilde{c} & \tilde{d}\end{pmatrix}=\pm 2$$

in the both cases $\mathrm{I_l}$ and $\mathrm{I_r}$.

5.2 We see that each couple

$$(a\,,b) \text{ with } a^2+b^2 \;=\; 3\,x \;\pm\; 2\,, \qquad (c\,,d) \text{ with } c^2+d^2 \;=\; 3\,x \;\mp\; 2\,, \qquad \text{or}$$

$$(r\,,s) \quad \text{with} \quad r^2+s^2 \;=\; x$$

restores the triple $(x\;>\;y\;>\;z)$. If there exists another triple $(x'\;>\;y'\;>\;z')$ with

$$x' \;=\; x\;,$$

then the corresponding semi-ambiguous forms A and A' differ by an ambiguous form B. On the other hand, we have two different decompositions into sum of two squares for x, $3x-2$, $3x+2$. They provide decompositions into coprime factors

$$x = x_1\,x_2\,, \quad 3\,x-2 = p_1\,p_2\,, \quad 3\,x+2 = q_1\,q_2 \quad \text{for odd } x \text{ and}$$

$$x = 2\,x_1\,x_2\,, \quad 3\,x-2 = 4\,p_1\,p_2\,, \quad 3\,x+2 = 8\,q_1\,q_2 \quad \text{for even } x\,,$$

where $x_1\,,\;x_2\,,\;p_1\,,\;p_2\,,\;q_1\,,\;q_2\;\geq\;5$. It is not hard to check that the ambiguous form B corresponds to the decompositions of $3x-2$, $3x+2$ and represents $\pm x_1^2$ and $\mp x_2^2$, so existence of a counterexample to the conjecture is equivalent to the solvability of the system

$$\begin{cases} r\,k^2 - s\,l^2 = \;\;\,4\,x_1^2 \\ r\,m^2 - s\,n^2 = -4\,x_2^2 \\ r\,s = 9\,x_1^2\,x_2^2 - 4 \end{cases}$$

for odd x, or the system

$$\begin{cases} r\,k^2 - 2\,s\,l^2 = \;\;\,x_1^2 \\ r\,m^2 - 2\,s\,n^2 = -\,x_2^2 \\ 8\,r\,s = 9\,x_1^2\,x_2^2 - 1 \end{cases}$$

for even x. Furthermore, for $x_2 > x_1$ it is easy to prove that $x_2 < x_1^3/9$.

Finally we note that the corresponding conjecture is true for the polynomial solutions in section 3.

References

[1] A.A.Markoff. *Sur les formes binaires indéfinies.* Math. Ann. **15, 17** (1879,1880), 381-409, 379-399.

[2] A.Hurwitz. *Uber eine Aufgabe der unbestimmten Analisis.* Archiv der Math. u. Phys. **3/11** (1907), 185-196.

[3] G.Frobenius. *Uber die Markoffschen Zahlen.* Preuss. Akad. Wiss. Sitzungsberichte. (1913), 458-487.

[4] Z.I.Borevich, I.R.Shafarevich. *Number Theory.* Academic Press Inc. (1966).

[5] D.Zagier. *On the Number of Markoff Numbers Below a Given Bound.* Math. Comp. **39** (1982), 709-723.

[6] A.L.Gorodentsev, A.N.Rudakov. *Exceptional vector bundles on projective spaces.* Duke Math.J. **54** (1987), 115-130.

[7] A.N.Rudakov. *The Markov numbers and exceptional bundles on* $\mathbb{P}^2$. English transl. in Math. SSSR Izv. **32** (1989), 99-112.

[8] J.H.Silverman. *The Markov Equation* $X^2+Y^2+Z^2=aXYZ$ *over Quadratic Imaginary Fields.* J. Number Theory. **35** (1990), 72-104