Advances in Applied Mathematics 127 (2021) 102179

Contents lists available at ScienceDirect

Advances in Applied Mathematics

www.elsevier.com/locate/yaama

Christoffel words and weak Markoff theory

Christophe Reutenauer

Département de mathématiques, Université du Québec à Montréal, Canada

ARTICLE INFO

Article history:
Received 21 November 2020
Received in revised form 28 December 2020
Accepted 1 January 2021
Available online xxxx

MSC:
05A05
68R15
11A55
11E16

Keywords:
Combinatorics on words
Christoffel words
Finite Sturmian words
Continued fractions
Indefinite binary quadratic forms

ABSTRACT

We call weak Markoff theory the theory of Markoff restricted to integral quadratic forms (instead of real ones), and to quadratic real numbers (instead of general real numbers). We show that weak Markoff theory may be reduced to combinatorial properties of Christoffel words, avoiding the use of bi-infinite sequences. These properties are two new characterizations of Christoffel words, one using the lexicographical order, the other being of arithmetic nature, based on values of continuant polynomials.

Contents

E-mail address: Reutenauer.Christophe@uqam.ca.

https://doi.org/10.1016/j.aam.2021.102179
0196-8858/

1. Introduction

In his 1875 article [5], Christoffel introduces the words that are now named after him; in 1876, Smith [22] rediscovers them. In his 1879 and 1880 articles [14,15], Markoff uses these words, likely not knowing the work of the two previous authors: he constructs certain quadratic forms and certain quadratic numbers (the latter by their continued fraction), which satisfy special inequalities.

Christoffel words, which are discrete objects, are interesting for themselves in several areas of pure and applied mathematics: combinatorics on words, Sturmian sequences, discrete geometry, continued fractions, free groups on two generators. See the following books: Lothaire [13], Aigner [1], and [19].

Coming back to Markoff theory, it has two parts: one on *real* quadratic forms (the actual work of Markoff), and one on approximation of *real* numbers. Both theories appear in several textbooks and articles: Hurwitz [12], Frobenius [10], Perron [16,17], Heawood [11], Remak [18], Bachmann [2], Dickson [8], Cassels [4], Cusick and Flahive [7], Bombieri [3], Aigner [1], the author [19].

The proofs in both Markoff theories are somewhat complicated. One has to construct bi-infinite sequences arising from continued fractions subject to certain inequalities: one must show that they are periodic, with a period equal to (roughly speaking) a Christoffel word.[1] The complexity of this proof is illustrated by the fact that it has been revisited many times over the past 140 years, as revealed by the previous list of authors.

In the present article, we present an alternative approach to both Markoff theories, which is somewhat simpler, at the cost of weakening the theories. The weakening is that we presuppose that the quadratic forms have integral coefficients, and that the real numbers are quadratic. Then we may reduce the most technical part of the proofs to properties of Christoffel words. This leads to two new characterizations of these words, which are interesting in themselves. One of them is of arithmetical nature, using continuant polynomials, and is a property of the values of these polynomials, when evaluated in sequences of 1 and 2's.

This property is of some independent interest, due to the fact that Markoff numbers are such values (Frobenius [10] (6) p. 612, see also [19] Theorem 10.3.5), so that the values

[1] However Cassels' approach is quite different.

of continuant polynomials on arguments 1, 2 are related to the Markoff numbers uniqueness conjecture (see the book of Martin Aigner [1] for many equivalent formulations of this conjecture, and solutions in many particular cases).

The present article has therefore two parts: the first part is on combinatorics on words, with two new characterizations of Christoffel words (more precisely of their conjugation classes). The first characterization, Theorem 4.1, uses the lexicographic ordering, and a property that we call *finitary Markoff property*, since it has some similarities with the property of bi-infinite sequences already considered by Markoff. The second characterization, Theorem 4.2, is of a new kind: it is arithmetic, and uses the Euler continuant polynomials, evaluated in the sequence of numbers which are the letters of Christoffel words and their conjugates (once the binary alphabet is replaced by 11 and 22).

The second part is applied to prove what we call *weak Markoff theory*. First, recall that Markoff theory has two aspects. One concerns real binary quadratic forms which satisfy some inequality relating their minimum and their discriminant; Markoff then shows that they must be very special, in the sense that they are, up to a real factor, $GL_2(\mathbb{Z})$-equivalent to a so-called *Markoff form*, which is one of a countable set of integral forms, parametrized by Markoff numbers. The other aspect[2] concerns approximation of real numbers, and more precisely reals whose *Lagrange number* is smaller than 3; then it is shown that these numbers are $GL_2(\mathbb{Z})$ to a so-called *Markoff irrationality*, which is one of a countable set of quadratic numbers, parametrized by Markoff numbers too.

In both theories, one has to prove combinatorial properties of infinite words, or bi-infinite words. What we call weak Markoff theory is the restriction of these Markoff's theories to integral forms on one hand, and on quadratic numbers on the other. In this case, one has to deal only with finite words, and actually only use the results of the first part of the article.

For the moment, I do not see how to deduce general Markoff theory from the weak one. To do this, one should prove the following rather strong results: (i) if the Lagrange number of a real number is < 3, then this number is quadratic; (ii) if a real indefinite binary quadratic form f satisfies $\frac{\sqrt{d(f)}}{m(f)} < 3$, then f is proportional to an integral form ($d(f)$ is the discriminant of f and $m(f)$ is the infimum of the values of f for integral values of x, y, not both 0). Of course, these results are true, by the general Markoff theory (see [19] for example), but the point would be to give simpler proofs.

Part 1. Words

2. Continuant polynomials

Continuant polynomials are defined for any $k \geq 0$ and any integers $n_1, \ldots, n_k$ as follows: $p_{-1} = 0, p_0 = 1$ and for any $k \geq 1$, $p_k(n_1, \ldots, n_k) = p_{k-1}(n_1, \ldots, n_{k-1})n_k +$

[2] This part of the theory, although attributed commonly to Markoff, is not stricto sensu in his articles, which are about quadratic forms. It was Hurwitz who first claimed that it could be proved by the same methods as Markoff's ([12] p. 284).

$p_{k-2}(n_1, \dots, n_{k-2})$ (*right recursion formula*). It is customary to drop the index k and to write $p(n_1, \dots, n_k)$ for $p_k(n_1, \dots, n_k)$. One has ([6] p. 116)

$$P(n_1) \cdots P(n_k) = \begin{pmatrix} p(n_1, \dots, n_k) & p(n_1, \dots, n_{k-1}) \\ p(n_2, \dots, n_k) & p(n_2, \dots, n_{k-1}) \end{pmatrix}, \tag{1}$$

where $P(n) = \begin{pmatrix} n & 1 \\ 1 & 0 \end{pmatrix}$. By associativity of the matrix product, one obtains the *left recursion formula*: $p(n_1, \dots, n_k) = n_1 p(n_2, \dots, n_k) + p(n_3, \dots, n_k)$. It follows also, by transposing the product, and using the symmetry of the matrices $P(a)$, that $p(n_1, \dots, n_k) = p(n_k, \dots, n_1)$.

We mention the *leapfrog construction* as it is stated in [6] p. 117: $p(n_1, \dots, n_k)$ is equal to the sum, without multiplicity, of $n_1 \cdots n_k$ and of all terms obtained by omitting in this term one or more pairs of adjacent factors $n_i n_{i+1}$. For example, $p(n) = n$, $p(n, m) = nm + 1$, $p(n, m, q) = nmq + n + q$, $p(n, m, q, r) = nmqr + nm + nr + qr + 1$.

We denote by $\mathbb{P}$ the set of positive (> 0) integers and by $\mathbb{P}^*$ the free monoid generated by $\mathbb{P}$. For each word u in $\mathbb{P}^*$, $p(u)$ denotes the continuant polynomial $p(n_1, \dots, n_k)$, where $u = n_1 \cdots n_k$ (n_i in $\mathbb{P}$). We denote also $P(u)$ for $P(n_1, \dots, n_k)$; thus P is here a monoid homomorphism from $\mathbb{P}^*$ into $GL_2(\mathbb{Z})$.

For a nonempty word $u = au'$, with first letter a, ${}^-u$ denotes the word u'. The notation u^- is defined symmetrically.

We order words of equal length in $\mathbb{P}^*$ by the *alternating lexicographical order*, defined recursively as follows: $u \leq_{alt} v$ if either u, v are both empty, or if $u = au', v = bv'$ ($a, b \in \mathbb{P}$) and either $a < b$, or $a = b$ and $u' \geq_{alt} v'$.

Lemma 2.1. *Suppose that u, v are words in $\mathbb{P}^*$ of equal positive length, such that $p(u) \geq p(v)$ and $p({}^-u) \leq p({}^-v)$. Then $u \geq_{alt} v$.*

It follows from the definition and the recursion formulas for continuant polynomials that for any word $w \in \mathbb{P}^*$, one has $p(w) > 0$; moreover, $p(w) \leq p(cw)$ if $c \in \mathbb{P}$. We use this in the proof below.

Proof. Let $u = ax$, $v = by$, $a, b \in \mathbb{P}$. Note that by hypothesis $p(x) \leq p(y)$.

If x, y are empty, then by hypothesis $u \geq v$ (as numbers) and therefore $u \geq_{alt} v$.

Suppose now that x, y are nonempty. If $a > b$, then $u \geq_{alt} v$.

If $a = b$, then by the left recursion formula for continuant polynomials, $ap(x) + p({}^-x) = p(u) \geq p(v) = ap(y) + p({}^-y)$, and $p(x) \leq p(y)$, so that we must have $p({}^-x) \geq p({}^-y)$. By induction on the length, we deduce that $x \leq_{alt} y$, hence $u \geq_{alt} v$.

If $a < b$, then $bp(y) + p({}^-y) = p(v) \leq p(u) = ap(x) + p({}^-x) \leq ap(x) + p(x) = (a+1)p(x) \leq bp(x) \leq bp(y)$; this is possible only if $p({}^-y) = 0$, which is not true. □

Note that this lemma may also be proved by using continued fractions.

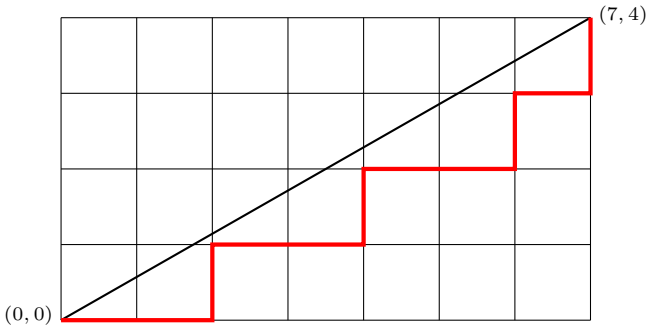

Fig. 1. The lower Christoffel word *aabaabaabab* of slope 4/7.

3. Christoffel words

The *lower Christoffel word of slope* $\frac{m}{l}$ is the word that discretizes from below a segment from $[0,0]$ to $[l,m]$ in the plane, as is shown in Fig. 1; here l, m are two relatively prime nonnegative integers. The *upper Christoffel word* is obtained by discretizing from above.

We denote by $\tilde{w}$ the reversal of the word w, obtained by writing the letters of w from right to left. A *palindrome* is a word equal to its reversal.

For each given slope, there is a lower Christoffel word w and an upper Christoffel word, which turns out to be the reversal $\widetilde{w}$.

The Christoffel words a and b have slope 1 or ∞ respectively. The other Christoffel words are called *proper*. See [1,19] for details.

4. Main theorems

We consider the alphabet a, b ordered by $a < b$ and we denote by $<_{lex}$ the lexicographic order on the free monoid $\{a < b\}^*$.

Two words are *conjugate* if they may be written xy and yx for some words x, y. Conjugation is an equivalence relation, and a corresponding equivalence class is called a *conjugation class*.

A word is *primitive* if it is not a nontrivial power of another word. A conjugation class is called primitive if each of its elements is primitive (equivalently, one of them).

The first theorem is a combinatorial characterization of *Christoffel classes* (= conjugation classes of Christoffel words).

Theorem 4.1. *Let C be a primitive conjugation class in $\{a, b\}^*$. The following conditions are equivalent:*

(i) *C is the set of the conjugates of some Christoffel word;*
(ii) *for each word in C of the form aub (resp. bua), one has $u \geq_{lex} \tilde{u}$ (resp. $u \leq_{lex} \tilde{u}$).*

Condition (ii) may be seen as a version for finite words of *Markoff's property* on bi-infinite words (see [19] Chapter 4): we call it the *finitary Markoff property*. Actually, it is equivalent to Markoff's condition for the bi-infinite word obtained by bi-infinite iteration of an element in the conjugation class C, and the previous theorem may be obtained as a consequence of a result of Markoff (see e.g. [3] Theorem 15, [1] Theorem 9.24, [19] Theorem 4.2.1). The point here however is to avoid infinite words, and give direct proofs.

We give now a numerical characterization of Christoffel classes. For $w \in \mathbb{P}^*$ of length at least 2, we set

$$\nu(w) = p(w) + p({}^-w^-) - 3p({}^-w);$$

and for $w = x$ of length 1, $\nu(w) = x - 3$.

We give two examples. Let $w \in \mathbb{P}^*$ be the word 1122, of length 4; then (see the formulas in Section 2)

$$\nu(1122) = p(1122) + p(12) - 3p(122) = p(1,1,2,2) + p(1,2) - 3p(1,2,2)$$

$$= (1\cdot 1\cdot 2\cdot 2 + 1\cdot 1 + 1\cdot 2 + 2\cdot 2 + 1) + (1\cdot 2 + 1) - 3\cdot(1\cdot 2\cdot 2 + 1 + 2) = 12 + 3 - 21 = -6.$$

Now let $v = 21$. Then

$$\nu(21) = p(2,1) + p() - 3p(1) = 3 + 1 - 3\cdot 1 = 1.$$

Define the monoid homomorphism χ from $\{a,b\}^*$ into $\{1,2\}^*$: it sends a onto 11 and b onto 22.

Theorem 4.2. *Let C be a primitive conjugation class in $\mathbb{P}^*$ of words of length at least 2. The following conditions are equivalent:*

(i) *For each word w in C, one has $\nu(w) \leq 0$, and $\nu(w) < 0$ if w is of odd length.*
(ii) *C is the conjugation class of $\chi(w)$ for some Christoffel word $w \in \{a,b\}^*$.*

The theorem is illustrated by the two previous examples: the word 1122 is the image under χ of the Christoffel word ab, and $\nu(1122) \leq 0$, and the reader may verify that the images under ν of all its conjugates are also ≤ 0; the word 21 is not in the image of χ and $\nu(12) > 0$.

For later use, note that for any nonempty word in $\mathbb{P}^*$

$$\nu(w) = tr(P(w)) - 3P(w)_{21}, \tag{2}$$

as follows from Eq. (1).

5. Proof of $(i) \Rightarrow (ii)$ in Theorem 4.1

We use a lemma which is of independent interest. Recall that a (finite) *Sturmian word* is a word in $\{a, b\}^*$ that is a factor of a Sturmian infinite word; equivalently it is a factor of some Christoffel word. It is known that a word w is Sturmian if and only if it is *balanced*, [13] Theorem 2.1.5; that is, for any factors u, v of equal length of w, the numbers of a's in u and v differ by at most 1.

Lemma 5.1. *A word w in $\{a, b\}^*$ is Sturmian if and only if for any factor of w of the form aub (resp. bua) one has $u \geq_{lex} \tilde{u}$ (resp. $u \leq_{lex} \tilde{u}$).*

Proof. If w is not a Sturmian word, then by a result of Morse and Hedlund (see [13] Proposition 2.1.3), w has the two factors ama and bmb for some palindrome m. Then either ama is at the left of bmb in w, or the contrary. In the first case, w has a factor aub such that ma is a prefix of u and bm is a suffix of u; then, $mb = \widetilde{m}b = \widetilde{bm}$ is a prefix of $\tilde{u}$ and therefore $u <_{lex} \tilde{u}$, and w does not satisfy the condition of the lemma. The second case is similar.

Suppose now that w is a Sturmian word. Let aub be a factor of w. Suppose by contradiction that $u <_{lex} \tilde{u}$. Then $u = pax, \tilde{u} = pby$ for some words p, x, y. Hence $u = \tilde{y}b\tilde{p}$ and $aub = apaxb = a\tilde{y}b\tilde{p}b$ has the factors apa and $b\tilde{p}b$ and is therefore not balanced, a contradiction. □

It is well-known that each conjugate of a Christoffel word is a Sturmian word (e.g. this follows from [19] Lemma 13.4.1). Hence the implication $(i) \Rightarrow (ii)$ in Theorem 4.1 follows from the lemma.

6. Proof of $(ii) \Rightarrow (i)$ in Theorem 4.1

This proof has many similarities with the proofs in Chapter 7 of [19]; however, it is not possible to simply refer to these proofs, because among other changes, the proof here is on finite words, and there on infinite words.

We assume here condition (ii) of Theorem 4.1. Below, when we use the letters x, y, it will mean that $\{x, y\} = \{a, b\}$.

1. We show first that no word of the form $xmxy\tilde{m}y$ is a factor of any word in C. Suppose the contrary. For example $amab\tilde{m}b$ is a factor of C. Then some word of C is of the form $b\tilde{m}bvama$. Then by (ii), we have $\tilde{m}bvam \leq_{lex} \tilde{m}a\tilde{v}bm$, a contradiction.

2. By 1. the following words are not a factor of C: $xxyy$, $xyxy^3$, $x(xy)^ixy(yx)^iy$, $xy(xy)^jxy(yx)^jyy$.

3. It follows directly from (ii) that no word of the form $xxvyy$ is in C.

4. We show that aa and bb are not both factor of any word in C. Suppose the contrary. Then $w = xxkyyt \in C$. We may assume that k is of minimum length. If k is empty, then $xxyy$ is a factor, contradicting 2. Thus, by minimality of the length of k, we have

$$w = xx(yx)^i yyt, i \geq 1, |k| = 2i.$$

If t is empty, then $w = xx(yx)^i yy \in C$, contradicting 3. Thus t is nonempty. It cannot begin by y, otherwise w contains the factor $x^2(yx)^i y^3$, hence also $xyxy^3$, contradicting 2. Thus t begins by x and $yt = (yx)^j s$, with $j > 0$, and chosen maximum, so that s does not begin by yx. Thus $w = xx(yx)^i y(yx)^j s = x(xy)^i xy(yx)^j s$.

If $j > i$, then w contains the factor $x(xy)^i xy(yx)^i y$, contradicting 2. Thus $j \leq i$.

If s is empty, then $w = x(xy)^i xy(yx)^j$ is conjugate to $(xy)^i xy(yx)^j x = (xy)^i xy^2(xy)^{j-1}x^2$, contradicting the minimality of $|k| = 2i$. Thus $|s| \geq 1$.

If s begins by x, then w has the factor $y(yx)^j x = y^2(xy)^{j-1}x^2$ and by minimality of $|k|$, we must have $j - 1 \geq i$, a contradiction with $j \leq i$.

If $s = y$, then $w = x(xy)^i xy(yx)^j y$, and, assuming $x = a, y = b$ (the other case is symmetric), we must have $(xy)^i xy(yx)^j \geq_{lex} (xy)^j yx(yx)^i$, a contradiction since $j \leq i$.

Thus s begins by y, but not by yx (by construction), and is not equal to y; hence s begins by yy, and w contains the factor $x(xy)^i xy(yx)^j y^2$. If $j = i$, then w contains the factor $x(xy)^i xy(yx)^i y$, contradicting 2. Thus $j \in \{1, \ldots, i-1\}$, and w contains $xy(xy)^j xy(yx)^j y^2$, contradicting 2. also.

5. Suppose that bb does not appear in C. We may assume that C is not the conjugation class of a, nor of b. Then we may find $w \in C$ such that w begins by a and ends by b. Then $w = G(u)$, for some shorter word u, where G is the endomorphism of the monoid $\{a, b\}^*$ sending a onto a and b onto ab. We show below that the conjugation class of u satisfies the finitary Markoff property, and we then may conclude by induction that C is the conjugation class of some Christoffel word, since G preserves Christoffel words (Corollary 2.6.2 in [19]).

Consider a conjugate of u of the form avb. Then $G(avb) = aG(v)ab$ is in C. Thus $G(v)a \geq_{lex} a\widetilde{G(v)} = G(\tilde{v})a$ (e.g. Lemma 4.1.3 in [19] for the last equality). Thus $G(v) \geq_{lex} G(\tilde{v})$ and therefore $v \geq_{lex} \tilde{v}$ since G is increasing for the lexicographical order (Lemma 3.13 in [21]).

Consider now a conjugate of u of the form bva. Then $G(bva) = abG(v)a$ is in C. This implies that $bG(v)aa \in C$. Hence $G(v)a \leq_{lex} a\widetilde{G(v)}$. Similarly as above, it implies $v \leq_{lex} \tilde{v}$.

6. If bb appears in C, then by 4. aa does not appear. Then we use the endomorphism $\tilde{D}$, sending a onto ab and b onto b, and conclude that C is the conjugation class of some Christoffel word.

7. Proof of $(ii) \Rightarrow (i)$ in Theorem 4.2

Define the monoid homomorphism $\mu : \{a, b\}^* \to SL_2(\mathbb{N})$ by $\mu(a) = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \mu(b) = \begin{pmatrix} 5 & 2 \\ 2 & 1 \end{pmatrix}$. Note that $P(1)^2 = \mu(a)$ and $P(2)^2 = \mu(b)$; hence for any word u in $\{a, b\}^*$

$$\mu(u) = P(\chi(u)). \tag{3}$$

Let w be a Christoffel word; then

$$\mu(w)_{12} = \frac{1}{3}tr(\mu(w)) \tag{4}$$

is the *Markoff number* associated with w (see e.g. [19] p. 38). It follows from Theorem 16.1 in [20] that for each conjugate u of w, one has $\mu(w)_{12} \leq \mu(u)_{12}$. Since the conjugation class of a Christoffel word is closed under reversal (see e.g. [19] Theorem 6.2.2), $\tilde{u}$ is also a conjugate of w, and thus

$$\mu(w)_{12} \leq \mu(\tilde{u})_{12} = \mu(u)_{21}, \tag{5}$$

since the transpose of $\mu(u)$ is $\mu(\tilde{u})$, the matrices $\mu(a)$ and $\mu(b)$ being symmetric.

Now, let v be a conjugate of $\chi(w)$. By the special form of the substitution χ, which maps a, b on $11, 22$ respectively, we have two cases:

(a) v is the image under χ of a conjugate u of w;
(b) $v = xmx$, $x = 1$ or 2, and mxx is the image under χ of a conjugate u of w.

In case (a), one has by Eq. (2) and Eq. (3): $\nu(v) = \mathrm{Tr}(P(v)) - 3P(v)_{21} = \mathrm{Tr}(P(\chi(u))) - 3P(\chi(u))_{21} = \mathrm{Tr}(\mu(u)) - 3\mu(u)_{21} = \mathrm{Tr}(\mu(w)) - 3\mu(u)_{21}$ (by conjugation) $= 3\mu(w)_{12} - 3\mu(u)_{21}$ (by Eq. (4)) ≤ 0, where the last inequality follows from Eq. (5).

In case (b), note first that $\chi(\tilde{u}) = \widetilde{\chi(u)} = \widetilde{mxx} = xx\tilde{m}$, and $\tilde{u}$ is a conjugate of w (since the conjugation class of w is closed under reversal); thus by (a), $\nu(xx\tilde{m}) \leq 0$. Next, $p(^{-}v) = p(mx) = p(x\tilde{m}) = p(^{-}(xx\tilde{m}))$. Hence, $P(v)_{21} = P(xx\tilde{m})_{21}$ by Eq. (1). Now, $\mathrm{Tr}(P(xx\tilde{m})) = \mathrm{Tr}(P(mxx))$ (since $P(xx\tilde{m})$ and $P(mxx)$ are transposed matrices) $= \mathrm{Tr}(P(xmx))$ (by conjugation) $= \mathrm{Tr}(P(v)$. Hence by Eq. (2), $\nu(v) = \mathrm{Tr}(P(v)) - 3P(v)_{21} = \mathrm{Tr}(P(xx\tilde{m})) - 3P(xx\tilde{m})_{21} = \nu(xx\tilde{m}) \leq 0$.

8. Proof of $(i) \Rightarrow (ii)$ in Theorem 4.2

We need two lemmas.

Lemma 8.1. *If $w = 21$, then $\nu(w) = 1$; if $w = 221^m$ or 2^m11 with m odd, then $\nu(w) = 0$.*

Proof. $\nu(21)$ was computed in Section 4.

Define the matrix function $\phi(A) = \mathrm{Tr}(A) - 3A_{21}$; then by Eq. (2), $\nu(w) = \phi(P(w))$; hence the sequence $a_k = \nu(221^{2k+1}) = \phi(P(221)(P(1)^2)^k)$ satisfies the linear recursion of length 2 determined by the characteristic polynomial of $P(1)^2$. But $a_0 = a_1 = 0$, because $P(221) = \begin{pmatrix} 7 & 5 \\ 3 & 2 \end{pmatrix}$ and $P(22111) = \begin{pmatrix} 19 & 12 \\ 8 & 5 \end{pmatrix}$. Thus the sequence vanishes and $\nu(221^m) = 0$.

For 2^m11, one argues similarly using the equalities $P(211) = \begin{pmatrix} 5 & 3 \\ 2 & 1 \end{pmatrix}$ and $P(22211) = \begin{pmatrix} 29 & 17 \\ 12 & 7 \end{pmatrix}$. □

Consider a word w having at least two distinct letters. A *circular block* in w is a factor $a_i a_{i+1} \cdots a_j$ of one of its conjugate $a_1 \cdots a_n$, where the a_k are letters, and $1 \leq i \leq j \leq n$, such that: $a_i = a_{i+1} = \ldots = a_j := a$ and where $a_{i-1} \neq a$ and $a_{j+1} \neq a$, where the indices $i-1, j+1$ are taken modulo n.

Lemma 8.2. *Suppose that C is a primitive conjugation class in $\{1,2\}^*$ such that if w or $\tilde{w} \in C$ and $w = 22u11$ (resp. $w = 11u22$), then $u \leq_{alt} \tilde{u}$ (resp. $u \geq_{alt} \tilde{u}$), where $\leq_{alt}$ is the alternating lexicographic order with $1 <_{alt} 2$. Assume that neither 121 nor 212 is a factor of any word in C, and that C does not contain any word of the form $1, 2, 21, 221^m$ or 2^m11 with m odd. Then the circular blocks of the words in C are of even length.*

Proof. 1. Suppose by contradiction that some word $w \in C$ has a circular block of 1's of odd length. Then some w in C is of the form either $w = 2^i1^m$, or $w = 2^i1^m2\cdots21^p$, with m odd, and $i, p \geq 1$ ($2\cdots2$ denotes a word beginning and ending by 2). Suppose that we are in the first case. If $m = 1$, we cannot have $i = 1$ by assumption; hence $i \geq 2$ and 212 is a factor of some conjugate of w, a contradiction. Thus $m \geq 3$, and we cannot have $i = 1$ (otherwise 121 is a factor of some conjugate); hence, by the hypothesis, $i \geq 3$, and thus $w = 222\cdots111$, which implies a contradiction, because $2\cdots1 >_{alt} 1\cdots2$.

2. We are therefore in the second case, and may thus assume that some word w in C is of the form $w = 2^i1^m2\cdots21^p$ with m odd and $i, p \geq 1$; we may assume that m is minimum possible. Since by hypothesis, 121 is not a factor in C, we must have $i \geq 2$. Since by hypothesis, 212 is not a factor in C, we must have $m \geq 3$ and $p \geq 2$. It follows from the hypothesis that 222111 is not a factor of w (otherwise w has a conjugate of the form $111\cdots222$, a contradiction); hence we must have $i = 2$. Thus $w = 221^m2\cdots21^p$ with m odd ≥ 3 and $p \geq 2$.

Suppose that p is odd; then by minimality of m, $p \geq m$. Assume first that $p = m$; then $w = 221^m2\cdots21^m$ and by the hypothesis, $1^m2\cdots21^{m-2} \leq_{alt} 1^{m-2}2\cdots21^m$. We may simplify this inequality at the left by the common prefix of even length 1^{m-3}, obtaining $1112\cdots \leq_{alt} 12\cdots$, a contradiction since the first letter is the same on both sides, so that by definition of the alternating lexicographical order, one should have $1 \geq 2$ (by considering the letters in rank 2). Thus we must have $p \geq m+2$. Next, w is conjugate to $1^m2\cdots21^p22$, whose reversal is $w' = 221^p2\cdots21^m$; by the hypothesis, we have $1^p2\cdots21^{m-2} \leq_{alt} 1^{m-2}2\cdots21^p$. Simplifying by 1^{m-3}, this implies $1^{p-m+3}\cdots \leq_{alt} 12\cdots$, a contradiction since $p - m + 3 \geq p - m \geq 2$.

Suppose now that p is even. With w' as above, we have by hypothesis $1^p2\cdots21^{m-2} \leq_{alt} 1^{m-2}2\cdots21^p$. If $p \geq m+1$, we simplify by 1^{m-3} and obtain $1^{p-m+3}\cdots \leq_{alt} 12\cdots$, a contradiction since $p - m + 3 \geq 2$. If $p \leq m-1$, we simplify by 1^{p-2} and obtain

$112\cdots \leq_{alt} 1^{m-p}2\cdots$. Assuming $p = m-1$, this gives $112\cdots \leq_{alt} 12\cdots$, a contradiction. If $p \leq m-3$, it gives a contradiction too, since $m-p \geq 3$.

3. For the blocks of 2's, it is enough to exchange 1 and 2, which replaces the alternating order by its opposite, noting that the forbidden factors in C, and the words in C, that appear in the hypothesis of the lemma, are stable under this exchange. □

Proof of $(i) \Rightarrow (ii)$ **in Theorem 4.2.** Suppose that each w in C satisfies $\nu(w) \leq 0$, with $\nu(w) < 0$ if w is of odd length.

1. Note first that $p(w) + p({}^{-}w^{-})$ is constant for w or $\tilde{w}$ in C. This is because this quantity is by Eq. (1) the trace of the matrix $P(a_0)\cdots P(a_{n-1})$, where $w = a_0\cdots a_{n-1}$, $a_i \in \mathbb{P}$, and because moreover, $p(w) = p(\tilde{w}), p({}^{-}w^{-}) = p({}^{-}\tilde{w}^{-})$.

2. We show that if w or $\tilde{w} \in C$, then $\nu(w) \leq 0$, with $\nu(w) < 0$ if w is of odd length. Since we know it by hypothesis for $w \in C$, it is enough to show that for each $w = ua \in C$ $(a \in \mathbb{P})$, we have $\nu(\tilde{w}) = \nu(au)$, noting that $au \in C$.

By assumption, w is of length at least 2. Then $\nu(\tilde{w}) = p(\tilde{w}) + p({}^{-}\tilde{w}^{-}) - 3p({}^{-}\tilde{w})$.

We have $\widetilde{{}^{-}\tilde{w}} = \widetilde{{}^{-}(a\tilde{u})} = \tilde{\tilde{u}} = u$. Therefore ${}^{-}(au) = u = \widetilde{{}^{-}\tilde{w}}$ and $p({}^{-}(au)) = p({}^{-}\tilde{w})$ (since p is invariant under reversal of the argument). Hence $\nu(\tilde{w})$ is equal by 1. to $p(au) + p({}^{-}(au)^{-}) - 3p({}^{-}(au)) = \nu(au)$.

3. We show now that C is on the alphabet $\{1,2\}$. Suppose the contrary. Then some w in C begins by $x \geq 3$. Since w is of length at least 2, $w = xyv$, $x, y \in \mathbb{P}$, and then $\nu(w) = xp(yv) + p(v) + p({}^{-}w^{-}) - 3p(yv) = (x-3)p(yv) + p(v) + p({}^{-}w^{-}) > 0$, a contradiction.

4. Note that if $w = xuy$, $x, y \in \mathbb{P}$, u nonempty, then by the recursive definition of p,

$$\nu(xuy) = (x-3)p(uy) + p({}^{-}uy) + p(u). \tag{6}$$

5. We show now that 121 is not a factor of any word in C. Suppose the contrary. Then some word w in C is of the form $w = 21v1$. Then by Eq. (6), $\nu(w) = -p(1v1) + p(v1) + p(1v)$. If v is empty, this is $-2+1+1 = 0$, a contradiction since w is of odd length. If $v = x$ is of length 1, then $\nu(w) = -(x+2) + x + 1 + x + 1 = x > 0$, a contradiction. Otherwise $v = amb$, $a, b \in \mathbb{P}$, $m \in \mathbb{P}^*$. It follows by the right and left recursion formulas of continuant polynomials that $\nu(w) = -p(1amb1) + p(amb1) + p(1amb) = -p(amb1) - p(mb1) + p(amb1) + p(1amb) = -p(mb1) + p(1amb) = -p(mb) - p(m) + p(amb) + p(mb) = p(amb) - p(m) > 0$ (since to each term in $p(m)$ corresponds a term, multiplied by ab, in $p(amb)$, and there are more terms in the latter), a contradiction again.

6. We show now that 212 is not a factor of any word in C. Suppose the contrary. Then some word w in C begins by 212. If w is of length 3, then $w = 212$ and thus by Eq. (6), $\nu(212) = -p(12) + p(2) + p(1) = 0$, a contradiction since w is of odd length. If w is of length ≥ 4, then w must begin by 2122, since by 5., 121 is not a factor. Then w has a conjugate u, which is in C, of the form $u = 22v21$; then by Eq. (6), $\nu(u) = \nu(22v21) = -p(2v21) + p(v21) + p(2v2) = -p(2v2) - p(2v) + p(v21) + p(2v2) = p(v21) - p(2v)$. If v is empty, this is $3 - 2 > 0$, a contradiction. If $v = a \in \mathbb{P}$, then

$\nu(u) = p(a21) - p(2a) = 2a + a + 1 - 2a - 1 = a > 0$, a contradiction. In general, $v = amb$, $a, b \in \mathbb{P}$, and $\nu(u) = p(amb21) - p(2amb) = p(amb2) + p(amb) - 2p(amb) - p(mb) = 2p(amb) + p(am) - p(amb) - p(mb) = p(amb) + p(am) - p(mb) > 0$, since $p(amb) - p(mb) \geq 0$, a contradiction.

7. We show that if $w = 22u11$ and if w or $\tilde{w} \in C$, then $u \leq_{alt} \tilde{u}$. We may assume that u is nonempty. Indeed, we have by Eq. (6), $\nu(w) = \nu(22u11) = -p(2u11) + p(u11) + p(2u1) = -p(2u1) - p(2u) + p(u11) + p(2u1) = -2p(u) - p(^-u) + p(u1) + p(u) = -p(u) - p(^-u) + p(u) + p(u^-) = -p(^-u) + p(u^-)$. This is ≤ 0, by 2., hence $p(u^-) \leq p(^-u)$. Thus $p(^-\tilde{u}) \leq p(^-u)$ (since p is invariant if one reverses its argument); since we have $p(u) = p(\tilde{u})$, it follows from Lemma 2.1 that $\tilde{u} \geq_{alt} u$.

8. By 2. and 7., we deduce that if $w = 11u22 \in C$, then $\tilde{u} \leq_{alt} u$.

9. From 5. and 6., we deduce that 121 and 212 are not factor of any word in C. From the hypothesis (i), Lemma 8.1 and 7. and 8., we deduce that the hypothesis of Lemma 8.2 is satisfied. Hence the circular blocks in C are of even length.

10. It follows from 9. that for some word u in $\{a, b\}$, C is the conjugation class of $\chi(u)$. Moreover, by 7. and 8., u satisfies the finitary Markoff property: indeed, for $x, y \in \{a, b\}^*$, $x <_{lex} y \Leftrightarrow \chi(x) <_{alt} \chi(y)$ (Lemma 5.6.3 in [19]). It follows then from Theorem 4.1 that u is conjugate to some Christoffel word and this proves (ii) in Theorem 4.2. □

Part 2. Weak Markoff theory

9. Preliminaries

Proposition 9.1. *Let* $\begin{pmatrix} a & b \\ c & d \end{pmatrix} = P(q_1) \dots P(q_n)$, $q_i \in \mathbb{P}$, $n \geq 1$. *Then the following inequalities are equivalent:*

(i) $\frac{(a+d)^2 - 4(-1)^n}{c^2} < 9$;
(ii) $a + d - 3c < 0$ *if* n *is odd, and* $a + d - 3c \leq 0$ *if* n *is even.*

Proof. Note that (i) is equivalent to

$$(a + d - 3c)(a + d + 3c) < 4(-1)^n. \tag{7}$$

1. Suppose that (i) holds. If n is odd, then $4(-1)^n < 0$, and $a + d + 3c > 0$, so that $a + d - 3c < 0$. If n is even, then $n \geq 2$, hence $a \geq 2, c, d \geq 1$ and $a + d + 3c \geq 6$; moreover, $4(-1)^n = 4$, so that Eq. (7) implies that $a + d - 3c \leq 0$.

2. Conversely, suppose that (ii) holds.

If n is odd, then $a + d - 3c \leq -1$ and we conclude that if $n > 1$, then a, d, c are positive, $a + d + 3c > 4$ and $(a + d - 3c)(a + d + 3c) < -4 = 4(-1)^n$; but if $n = 1$, then we must have $d = 0, c = b = 1$ and then, since $a + d < 3c$, we have $a = 1$ or 2 and $(a + d)^2 - 4(-1)^n = a^2 + 4 \leq 8 < 9 = 9c^2$.

If n is even, $a+d-3c \leq 0$ and $a+d+3c > 0$, so that $(a+d-3c)(a+d+3c) \leq 0 < 4 = 4(-1)^n$. □

Corollary 9.1. *Let $n \geq 2$, $q_1 \cdots q_n$ be a primitive word in $\mathbb{P}^*$ and for $i = 1, \ldots, n$, define the matrices $M_i = \begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix} = P(q_i) \cdots P(q_n)P(q_1) \ldots P(q_{i-1})$. Suppose that $\frac{(a_i+d_i)^2-4(-1)^n}{c_i^2} < 9$ for any $i = 1, \ldots, n$. Then for some i, the word $q_i \cdots q_n q_1 \cdots q_{i-1}$ is equal to $\chi(w)$ for some Christoffel word w.*

The case where $n = 1$ (which we do not need) may be settled directly: one has then $a_1 = 1$ or 2.

Proof. Let $w_i = q_i \cdots q_n q_1 \cdots q_{i-1} \in \mathbb{P}^*$. The w_i's are the conjugates of w_1. It follows from Eq. (1) that $p(w_i) = a_i, p(^-w_i) = c_i, p(^-w_i^-) = d_i$. Thus $\nu(w_i) = a_i + d_i - 3c_i$. By the hypothesis and the previous proposition, we deduce that $\nu(w_i) \leq 0$, with strict inequality if n is odd. Thus we conclude using Theorem 4.2. □

10. Approximation of reals

With each Christoffel word w associate the *Markoff irrationality* x_w, whose expansion into continued fractions is $[\overline{a_0, \ldots, a_{n-1}}]$, with $\chi(\tilde{w}) = a_0 \cdots a_{n-1}$; equivalently, $x_w = \frac{p-s+\sqrt{9q^2-4}}{2q}$, where $\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \mu(w) = {}^t\mu(\tilde{w})$, and where μ is the monoid homomorphism defined in Section 7 (see [19] p. 88). Since the words w and $\tilde{w}$ are conjugate, it follows that x_w is $GL_2(\mathbb{Z})$-equivalent to the quadratic number whose expansion into continued fractions is $\chi(w)$ repeated infinitely many times.

Recall that the *Lagrange number* of a real number ξ is the supremum of the positive reals C such that ξ has infinitely many rational approximations p/q such that $|\xi - p/q| < 1/Cq^2$.

Theorem 10.1 *(Markoff). Suppose that the Lagrange number of some quadratic real number ξ is < 3. Then ξ is $GL_2(\mathbb{Z})$-equivalent to some Markoff irrationality x_w.*

We may assume that ξ is *reduced*, that is, its expansion into continued fractions is periodic: $\xi = [\overline{q_1, \ldots, q_n}]$. We may assume that $n \geq 2$. Then by [1] Proposition 1.29, its Lagrange number is the maximum of the n numbers $\xi_i - \bar{\xi}_i$, $i = 1, \ldots, n$, where $\xi_i = [\overline{q_i, \ldots, q_n, q_1, \ldots, q_{i-1}}]$, and where the bar indicates quadratic conjugation.

We claim that $\xi_i - \bar{\xi}_i$ is equal to $\frac{\sqrt{(a_i+d_i)^2-4(-1)^n}}{c_i}$, with the notations of Corollary 9.1.

Admitting the claim, the hypothesis implies that $\frac{(a_i+d_i)^2-4(-1)^n}{c_i^2} < 9$. It follows then from Corollary 9.1 that for some i the word $q_i \cdots q_n q_1 \cdots q_{i-1}$ in $\mathbb{P}^*$ is equal to $\chi(w)$ for some Christoffel word w; this implies that ξ_i and x_w are $GL_2(\mathbb{Z})$-equivalent (see the sentence before the theorem), and this proves the theorem.

To prove the claim, it will suffice to prove it for $i = 1$. Note that $\xi_1 = \xi$ is the positive solution of the equation $\xi = \frac{a\xi+b}{c\xi+d}$, $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = P(q_1)\cdots P(q_n)$. Then $c\xi^2 + (d-a)\xi - b = 0$. Note that $\xi > 0$ and $\bar{\xi} < 0$ (because $\xi = [\overline{q_1, \ldots, q_n}]$ and $\xi\bar{\xi} = -b/c < 0$). It follows that the difference of the two roots is $\xi - \bar{\xi} = \frac{\sqrt{\Delta}}{c}$, where Δ is the discriminant of this equation. Next $\Delta = (d-a)^2 + 4bc = (a+d)^2 - 4ad + 4bc$ and the claim is proved.

11. Quadratic forms

Let $f(x, y) = ax^2 + bxy + cy^2$ be an indefinite binary quadratic form with integral coefficients; its discriminant, which is positive, is denoted $d(f)$. It is assumed that if $f(p, q) = 0$ for some integers p, q, then $p = q = 0$. Let $m(f) = \min\{|f(p, q)|, p, q \in \mathbb{Z}, (p, q) \neq (0, 0\}$.

With each Christoffel word w, we associate the Markoff form $f_w = qx^2 + (s-p)xy - ry^2$, where $\begin{pmatrix} p & q \\ r & s \end{pmatrix} = \mu(w)$, and where μ is as in Section 7 (see [19] Section 9.3).

Recall that, with f as above, each form $f(ax + by, cx + dy)$ with $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2(\mathbb{Z})$, is said to be $GL_2(\mathbb{Z})$-*equivalent* to f. Note that $d(f)$ and $m(f)$ are invariant under $GL_2(\mathbb{Z})$-equivalence.

Theorem 11.1 *(Markoff). Assume that f is an integral indefinite binary quadratic form, such that $\frac{\sqrt{d(f)}}{m(f)} < 3$. Then f is $GL_2(\mathbb{Z})$-equivalent to a form which is proportional to a Markoff form f_w.*

Call *roots* of the quadratic form $f(x, y)$ the roots of the quadratic polynomial $f(x, 1)$. We may assume that one root is ξ with $\xi > 1$ and the other $\bar{\xi} \in (0, 1)$. Indeed, in the $SL_2(\mathbb{Z})$-class of any form, there is a reduced one, which has by definition roots of opposite sign ξ and $\bar{\xi}$ satisfying $|\xi| > 1, |\bar{\xi}| < 1$ (see [9] definition on p. 100 and Theorem 76); then either $\xi > 1$ and then $\bar{\xi} \in (0, 1)$, or $\xi < -1$, and then we exchange x and y, and we are done.

Then, by Galois theorem (see [1] Proposition 1.18), the expansion of ξ in continued fractions is periodic: $\xi = [\overline{q_1, \ldots, q_n}]$. Let $M = P(q_1)\cdots P(q_n) = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ and $f_M(x, y) = cx^2 + (d-a)xy - by^2$. By the theory of continued fractions, ξ is fixed by the Möbius function associated to M, that is, $\xi = \frac{a\xi+b}{c\xi+d}$. Thus ξ is a root of the polynomial $f_M(x, 1)$ and it follows that f and f_M are proportional. Note that the discriminant of f_M is $(a+d)^2 - 4(-1)^n$.

Moreover, by Theorem 9.2.1 in [19], $|m(f)|$ is equal to the minimum of the first coefficients c_i of the quadratic forms f_{M_i}, with $M_i = P(q_i)\cdots P(q_n)P(q_1)\cdots P(q_{i-1}) =$

$\begin{pmatrix} a_i & b_i \\ c_i & d_i \end{pmatrix}$. Thus $\frac{(a_i+d_i)^2-4(-1)^n}{c_i^2} < 9$. Then Corollary 9.1 implies that for some i the word $q_i \cdots q_n q_1 \cdots q_{i-1}$ is equal to $\chi(w)$ for some Christoffel word w. This implies $M_i = P(\chi(w)) = \mu(w)$ (by Eq. (3)) and therefore, by definition of f_w, $f_w = b_i x^2 + (d_i - a_i)xy - c_i y^2$. Since $f_{M_i} = c_i x^2 + (d_i - a_i)xy - b_i y^2$, we have $f_{M_i}(x, y) = -f_w(-y, x)$. Finally, $f_M = f_{M_1}$ is $GL_2(\mathbb{Z})$-equivalent to $\pm f_w$ (Lemma 9.1 in [20]).

Acknowledgments

This work was partially supported by NSERC Canada.

References

[1] M. Aigner, Markov's Theorem and 100 Years of the Uniqueness Conjecture, Springer Verlag, 2013.
[2] P. Bachmann, Die Arithmetik der quadratischen Formen, Teubner, Leipzig, Berlin, 1923.
[3] E. Bombieri, Continued fractions and the Markoff tree, Expo. Math. 25 (2007) 187–213.
[4] J.W.S. Cassels, An Introduction to Diophantine Approximation, Cambridge University Press, 1957.
[5] E.B. Christoffel, Observatio arithmetica, Ann. Mat. Pura Appl. 6 (1875) 145–152.
[6] P.M. Cohn, Free Rings and Their Relations, Academic Press, 1985.
[7] T.W. Cusick, M.E. Flahive, The Markoff and Lagrange Spectra, AMS, 1989.
[8] L.E. Dickson, Studies in the Theory of Numbers, Chelsea, New York, 1957; first edition, 1930.
[9] L.E. Dickson, Introduction to the Theory of Numbers, Dover, 1957.
[10] G. Frobenius, Über die Markoffschen Zahlen, Sitz.ber. K. Preuss. Akad. Wiss. Berl. 26 (1913) 458–487.
[11] P.J. Heawood, The classification of rational approximations, Proc. Lond. Math. Soc. 20 (1922) 233–250.
[12] A. Hurwitz, Ueber die angenäherte Darstellung der Irrationalzahlen durch rationale Brüche, Math. Ann. 39 (1891) 279–284.
[13] M. Lothaire, Algebraic Combinatorics on Words, Cambridge University Press, 2002.
[14] A.A. Markoff, Sur les formes quadratiques binaires indéfinies, Math. Ann. 15 (1879) 381–496.
[15] A.A. Markoff, Sur les formes quadratiques binaires indéfinies (second mémoire), Math. Ann. 17 (1880) 379–399.
[16] O. Perron, Über die Approximationen irrationaler Zahlen durch rationale, Sitzungsber. Heidelb. Akad. Wiss. 4 (1921) 2–17.
[17] O. Perron, Über die Approximationen irrationaler Zahlen durch rationale II, Sitzungsber. Heidelb. Akad. Wiss. 8 (1921) 2–12.
[18] R. Remak, Über indefinite binäre quadratische Minimalformen, Math. Ann. 92 (1924) 155–182.
[19] C. Reutenauer, From Christoffel Words to Markoff Numbers, Oxford University Press, 2019.
[20] C. Reutenauer, On quadratic numbers and forms, and Markoff theory, submitted for publication.
[21] G. Richomme, Lyndon morphisms, Bull. Belg. Math. Soc. Simon Stevin 10 (2003) 761–785.
[22] H.J.S. Smith, Note on continued fractions, Messenger Math. 6 (1876) 1–14.