ON PATTERNS OCCURRING IN BINARY ALGEBRAIC NUMBERS

B. ADAMCZEWSKI AND N. RAMPERSAD

ABSTRACT. We prove that every algebraic number contains infinitely many occurrences of 7/3-powers in its binary expansion. Using the same approach, we also show that every algebraic number contains either infinitely many occurrences of squares or infinitely many occurrences of one of the blocks 010 or 02120 in its ternary expansion.

1. INTRODUCTION AND RESULT

The binary expansion of an algebraic irrational number such as

$$\sqrt{2} = 1.0110101000001001111001100110011111110011110111110011\cdots$$

is a source of a number of intriguing and challenging problems. In some respects, this topic was initiated by a famous note of É. Borel [4] in 1950. After this pioneering work, it was commonly believed that every algebraic irrational number is—as is almost every number with respect to the Lebesgue measure—a normal number. This belief, if true, would of course imply that every block of digits, or equivalently, every finite word in the free monoid $\mathcal{A} = \{0, 1\}^*$, would occur infinitely often in their binary expansions. Despite some recent progress on lower bounds for the number $p(n)$ of such blocks of a given length n (see [1]), this problem has hardly been tackled. Regarding occurrences of particular blocks of digits, the situation is even worse. The only known result is somewhat trivial: the words 0, 1, 01 and 10 have to appear infinitely often in the binary expansion of an irrational number. This is actually a straightforward consequence of the fact that an irrational number cannot have an eventually periodic expansion. However, if we fix an algebraic irrational number α and a (non-empty) finite word W in $\{0, 1\}^* \setminus \{0, 1, 01, 10\}$, it is still an open problem to determine whether the block of digits W occurs infinitely often or not in the binary expansion of α.

Instead of asking for occurrences of specific blocks of digits, we are interested here in what can be regarded as the combinatorial skeleton of binary expansions of algebraic numbers, in the sense that we will only consider patterns. Let us give an example of what we mean by pattern. A *square* is a pattern of the form XX, where $X \in \mathcal{A}^+$ (i.e., X non-empty). For instance, the words 00, 11 and 011 011 are different examples of squares, but all these words reflect the same pattern XX. Since every binary word of length 4 contains a square, such a pattern is unavoidable by binary infinite words, that is, every infinite binary word contains infinitely many occurrences of squares. Consequently, this is also the case for the binary expansion of every algebraic number, though we are not able to give any explicit example of such a square. If we consider some more repetitive patterns, things become

less obvious. Recall that an *overlap* is a pattern of the form $xXxXx$ with $x \in \mathcal{A}$ and $X \in \mathcal{A}^*$. In 1912, Thue [9] proved the existence of infinite binary words avoiding overlaps. More precisely, Thue suggested the following explicit example, now referred to as the Thue–Morse sequence (see the survey paper of Allouche and Shallit [3]). The Thue–Morse sequence $\mathbf{t} = (t_n)_{n \geq 0}$ on the alphabet $\{0, 1\}$ is defined as follows: $t_n = 0$ (respectively, $t_n = 1$) if the sum of the binary digits of n is even (respectively, odd). An interesting property of the sequence $\mathbf{t}$ is that it is the fixed point beginning with 0 of the binary morphism σ defined by $\sigma(0) = 01$ and $\sigma(1) = 10$. This morphism is usually called the *Thue–Morse morphism*.

Before stating our result, we need to introduce some notation. The length of a finite word W on the alphabet $\mathcal{A}$, that is, the number of letters composing W, is denoted by $|W|$. For any positive integer k, we write W^k for the word $W \cdots W$ (the word W concatenated k times). More generally, for any positive real number α, we denote by W^α the word $W^{\lfloor \alpha \rfloor} W'$, where W' is the prefix of W of length $\lceil (\alpha - \lfloor \alpha \rfloor)|W| \rceil$. Here, $\lfloor y \rfloor$ and $\lceil y \rceil$ denote, respectively, the integer part and the upper integer part of the real number y. Such a word is called a α-*power*. For instance, the word 011 011 011 is a 3-power or a cube, while the word 011010 011010 01 is a 7/3-power.

As a consequence of the expected normality of algebraic irrational numbers, every repetitive pattern X^α, with $\alpha > 1$, should occur infinitely often in their binary expansions. In this note we prove the following result, providing a very small step in this direction. Nevertheless, it seems to be the first result of this type.

Theorem 1.1. *The binary expansion of an algebraic number contains infinitely many occurrences of 7/3-powers.*

In 1929, Mahler proved that the binary Thue–Morse number

$$\sum_{n \geq 0} \frac{t_n}{2^n}$$

is transcendental. As a generalization of Mahler's result, we quote the following straightforward consequence of Theorem 1.1.

Corollary 1.2. *The binary expansion of an algebraic number contains infinitely many occurrences of overlaps.*

Our proof of Theorem 1.1 relies on two recent results: on one side, a combinatorial transcendence criterion due to Adamczewski, Bugeaud and Luca [2], which is based on the p-adic version of the Schmidt subspace theorem, and, on the other side, a structure theorem for binary words avoiding 7/3-powers obtained by Karhumäki and Shallit [5]. In fact, we will see that, quite surprisingly, infinite words avoiding 7/3-powers turn out to be rather repetitive in a different setting. This new property gives rise to the transcendence of the associated binary numbers.

2. PROOF OF THEOREM 1.1

Let $\mathbf{a} = (a_k)_{k \geq 1}$ be a sequence of elements from $\mathcal{A}$, which we identify with the infinite word $a_1 a_2 \cdots$. Let $w > 1$ be a real number. We say that $\mathbf{a}$

satisfies Condition $(*)_w$ if there exist two sequences of finite words $(U_n)_{n \geq 1}$, $(V_n)_{n \geq 1}$ such that:

(i) For any $n \geq 1$, the word $U_n V_n^w$ is a prefix of the word $\mathbf{a}$;

(ii) The sequence $(|U_n|/|V_n|)_{n \geq 1}$ is bounded from above;

(iii) The sequence $(|V_n|)_{n \geq 1}$ is increasing.

A sequence satisfying Condition $(*)_w$ for some $w > 1$ is called a *stammering sequence* or a *stammering word*.

In [2], Adamczewski, Bugeaud and Luca proved the following result regarding stammering real numbers.

Theorem ABL. *Let $b > 1$ be an integer and let $\mathbf{a} = (a_n)_{n \geq 0}$ be a stammering sequence over $\{0, 1, \ldots, b - 1\}$. Then, the real number*

$$\sum_{n \geq 0} \frac{a_n}{b^n}$$

is either rational or transcendental.

The second main ingredient in the proof of Theorem 1.1 is the following structure theorem of Karhumäki and Shallit [5]. We mention that in the case where 7/3-powers are replaced by overlaps, this result was proved by Restivo and Salemi [8] .

Theorem KS. *Let $\mathbf{x}$ be an infinite binary word avoiding 7/3-powers and let σ denote the Thue–Morse morphism. Then there exist $u \in \{\epsilon, 0, 1, 00, 11\}$ and an infinite binary word $\mathbf{y}$ avoiding 7/3-powers such that $\mathbf{x} = u\sigma(\mathbf{y})$.*

We are now ready to deduce Theorem 1.1 from a simple combination of Theorem ABL and Theorem KS.

Proof of Theorem 1.1. First, since rational numbers have eventually periodic binary expansions, these expansions contain arbitrarily large powers and *a fortiori* infinitely many occurrences of 7/3-powers. We thus have only to consider algebraic irrational numbers. In virtue of Theorem ABL, it remains to prove that infinite binary words avoiding 7/3-powers are stammering words. We will deduce the latter result from Theorem KS.

Let $\mathbf{a} = (a_n)_{n \geq 0}$ be a binary sequence avoiding 7/3-powers. Let $k \geq 1$ be an integer. Applying Theorem KS k times to the infinite word $\mathbf{a}$, we obtain the following factorization:

$$\mathbf{a} = u_1 \sigma(u_2) \sigma^2(u_3) \cdots \sigma^{k-1}(u_k) \sigma^k(\mathbf{y}'),$$

where each u_i has length at most 2 and $\mathbf{y}'$ is an infinite word avoiding 7/3-powers. Recall that every binary word of length 4 contains a square. There thus exists a possibly empty word A of length at most 2 and a word B of length 1 or 2 such that $\mathbf{y}'$ begins with ABB. Set

$$U_k = u_1 \sigma(u_2) \sigma^2(u_3) \cdots \sigma^{k-1}(u_k) \sigma^k(A)$$

and

$$V_k = \sigma^k(B).$$

Then, the sequence $\mathbf{a}$ begins with $U_k V_k^2$. Moreover, we infer from a simple computation that

$$|U_k| < 2^{k+2} \quad \text{and} \quad 2^k \leq |V_k| \leq 2^{k+1}.$$

Without loss of generality, we can assume that the sequence $(|V_n|)_{n \geq 1}$ increases (indeed, we can, if necessary, consider the increasing subsequence $(V_{2n})_{n \geq 1}$). Consequently, $\mathbf{a}$ satisfies Condition $(*)_2$ and is a stammering sequence. This proves that the binary expansion of every algebraic number contains at least one occurrence of a 7/3-power. Since algebraicity does not depend on the first digits of the binary expansion, this implies that every algebraic number contains infinitely many occurrences of 7/3-powers, concluding the proof. $\qquad \square$

3. COMMENTS AND OPEN PROBLEMS

Questions about patterns occurring in expansions of algebraic numbers are of course not restricted to the case of the base 2. However, even with ternary expansions, new difficulties appear. For instance, there exist infinite ternary words avoiding squares, so that we do not know whether the ternary expansion of an algebraic number always contains infinitely many occurrences of squares. In this direction, we remark that the approach used for proving Theorem 1.1 leads to the following weaker result.

Theorem 3.1. *The ternary expansion of an algebraic number contains either infinitely many occurrences of squares or infinitely many occurrences of one of the blocks* 010 *or* 02120.

Proof. Every rational number contains infinitely many squares in its ternary expansion, so it suffices to consider algebraic irrational numbers. To prove the desired result it is enough to show that any ternary word that avoids squares as well as the subwords 010 and 02120 is a stammering word.

Let $\mathbf{a} = (a_n)_{n \geq 0}$ be a ternary sequence avoiding squares as well as 010 and 02120. Define a morphism τ by $\tau(0) = 011$, $\tau(1) = 01$, and $\tau(2) = 0$. We need the following result due to Thue [9] (see Problem 2.3.7 of Lothaire [6]): a word $\mathbf{a}$ avoids squares and the subwords 010 and 02120 if and only if $\mathbf{b} = \tau(\mathbf{a})$ is overlapfree.

Consider the infinite binary word $\mathbf{b}$. It is overlapfree and thus *a fortiori* 7/3-powerfree. We may therefore apply the previous argument to show that for every integer $k \geq 1$, $\mathbf{b}$ begins with a prefix $U_k V_k^2$, where $|U_k| \leq 2^{k+2}$ and $|V_k| \geq 2^k$.

Observe that for any overlapfree binary word, between any two successive occurrences of 0 there can be at most two 1's; otherwise, we would have an occurrence of the overlap 111. By considering the number of 1's between every two successive occurrences of 0, we see that the word $\mathbf{b}$ has a unique factorization into the blocks 011, 01, and 0.

Let x and y be the shortest words such that $x V_k^2 y$ is a subword of $\mathbf{b}$ and $x V_k^2 y$ begins and ends with 0. Now write

$$x V_k^2 y = 01^{i_1} 01^{i_2} \cdots 01^{i_\ell} 0,$$

where $i_j \in \{0,1,2\}$ for $j = 1, \dots, \ell$. Set $W_k = w_1 w_2 \dots w_\ell$, where $w_j = (2 - i_j)$ for $1 \le j \le \ell$. Then, W_k is a subword of $\mathbf{a}$ and has either the form $XaXb$ or $bXaX$, where $a, b \in \{0,1,2\}$ and $a \ne b$.

It follows that for every $k \ge 1$, there exist words Y_k and Z_k, and a letter $a \in \{0,1,2\}$, such that $\mathbf{a}$ begins with $Y_k Z_k a Z_k$, where

$$|Y_k| \le 2^{k+2} \text{ and } |Z_k| \ge \frac{2^k}{3}.$$

Thus $\mathbf{a}$ satisfies Condition $(*)_w$ for every fixed w in the interval $(1, 2)$ provided that $k > k_w$ is sufficiently large. Consequently, $\mathbf{a}$ is a stammering sequence. This concludes the proof. □

We end this note with the two following open questions regarding the binary expansion of algebraic numbers.

Question 2.1. *Is it true that the binary expansion of every algebraic number contains arbitrarily large squares?*

Given a finite word $W = w_1 w_2 \cdots w_r$, the reversal of W, denoted by W^R, is the word $w_r w_{r-1} \cdots w_1$. A palindrome is a pattern invariant by reversal. It is either of the form XX^R if its length is even, or of the form XxX^R if its length is odd.

Question 2.2. *Is it true that the binary expansion of every algebraic number contains arbitrarily large palindromes?*

Not surprisingly, a positive answer is expected for both questions.

Acknowledgements

References

[1] B. Adamczewski & Y. Bugeaud, On the complexity of algebraic numbers I. Expansion in integer bases, *Annals of Math.* **165** (2007), 547–565.

[2] B. Adamczewski, Y. Bugeaud & F. Luca, Sur la complexité des nombres algébriques, *C. R. Acad. Sci. Paris* **339** (2004), 11–14.

[3] J.-P. Allouche & J. Shallit, The ubiquitous Prouhet-Thue-Morse sequence. In C. Ding. T. Helleseth, and H. Niederreiter, eds., *Sequences and Their Applications: Proceedings of SETA '98*, Springer-Verlag, 1999, pp. 1–16.

[4] É. Borel, Sur les chiffres décimaux de $\sqrt{2}$ et divers problèmes de probabilités en chaîne, *C. R. Acad. Sci. Paris* **230** (1950), 591–593.

[5] J. Karhumäki & J. Shallit, Polynomial versus exponential growth in repetition-free binary words, *J. Combin. Theory, Ser. A* **105** (2004), 335–347.

[6] M. Lothaire, *Combinatorics on Words*, Vol. 17 of Encyclopedia of Mathematics and its Applications, Addison-Wesley, 1983.

[7] K. Mahler, Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen, *Math. Annalen* **101** (1929), 342–366. Corrigendum **103** (1930), 532.

[8] A. Restivo & S. Salemi, "Overlap free words on two symbols". In M. Nivat, D. Perrin, eds., *Automata on Infinite Words*, Vol. 192 of *Lecture Notes in Computer Science*, pp. 198–206, Springer-Verlag, 1984.

[9] A. Thue, Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, *Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912) 1–67; reprinted in *Selected Mathematical Papers of Axel Thue*, T. Nagell, ed., Universitetsforlaget, Oslo, 1977, pp. 413–478.

Boris Adamczewski, CNRS, Université Lyon 1, Université de Lyon, Institut Camille Jordan, 21 avenue Claude Bernard, 69622 Villeurbanne cedex France

Narad Rampersad, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1