Fizikos ir matematikos fakulteto
Seminaro darbai,
Šiaulių universitetas,
8, 2005, 5–13

ON THE DECIMAL EXPANSION OF ALGEBRAIC NUMBERS

Boris ADAMCZEWSKI[1], **Yann BUGEAUD**[2]

[1] *CNRS, Institut Camille Jordan, Université Claude Bernard Lyon 1, Bât. Braconnier, 21 avenue Claude Bernard, 69622 Villeurbanne Cedex, France; e-mail: Boris.Adamczewski@math.univ-lyon1.fr*

[2] *Université Louis Pasteur, U. F. R. de mathématiques, 7, rue René Descartes, 67084 Strasbourg Cedex, France; e-mail: bugeaud@math.u-strasbg.fr*

Abstract. In this expository paper, we discuss various combinatorial criteria that may apply to the decimal (or, more generally, to the b-adic) expansion of a given real number to show that this number is transcendental. As a consequence, we show that the sequence of decimals of $\sqrt{2}$ cannot be "too simple".

Key words and phrases: b-adic expansion, integer base, Fibonacci word.

Mathematics Subject Classification: 11J81, 11A63, 11B85, 11K16.

1. Introduction

Throughout the present paper, b always denotes an integer $\geqslant 2$ and ξ is a real number with $0 < \xi < 1$. There exists a unique infinite sequence $\mathbf{a} = (a_j)_{j \geqslant 1}$ of integers in $\{0, 1, \ldots, b-1\}$, called the b-adic expansion of ξ, such that

$$\xi = \sum_{j \geqslant 1} \frac{a_j}{b^j},$$

and $\mathbf{a}$ does not terminate in an infinite string of 0. Clearly, the sequence $\mathbf{a}$ is ultimately periodic if, and only if, ξ is rational. With a slight abuse of notation, we also call $\mathbf{a}$ the infinite word $\mathbf{a} = a_1 a_2 \ldots$. For our purpose, it is much more convenient to use the terminology of combinatorics on words rather than working with sequences. Obviously, $\mathbf{a}$ depends on ξ and b, but we choose not to indicate this dependence: this notation will be kept throughout the paper.

Recall that the real number ξ is called *normal in base b* if, for any positive integer n, each one of the b^n words of length n on the alphabet $\{0, 1, \ldots, b-1\}$

occurs in the b-adic expansion of ξ with the same frequency $\frac{1}{b^n}$. The first explicit example of a number normal in base 10, namely the number

$$0.1234567891011121314\ldots, \tag{1}$$

whose sequence of digits is the concatenation of the sequence of all positive integers ranged in increasing order, was given in 1933 by Champernowne [6]. It follows from the Borel-Cantelli lemma that almost all real numbers (in the sense of the Lebesgue measure) are normal in every base b, but proving that a specific number, like e, π or

$$\sqrt{2} = 1.4142135623730950488016887\ldots,$$

is normal in some base remains a challenging open problem. However, the following conjecture is widely believed to be true (recall that an algebraic number is a root of a non-zero polynomial with integer coefficients).

Conjecture 1. *Every real irrational algebraic number is normal in every integer base.*

Conjecture 1 is reputed to be out of reach, thus, we will focus our attention to simpler questions. A natural way to measure the *complexity* of ξ (in base b) is to count the number of distinct blocks of given length in the infinite word $\mathbf{a}$ defined above. Thus, for an infinite word $\mathbf{w}$ on the alphabet $\{0, 1, \ldots, b-1\}$ and for any positive integer n, we let $p(n, \mathbf{w})$ denote the number of distinct blocks of n letters occurring in $\mathbf{w}$. Furthermore, we set $p(n, \xi, b) = p(n, \mathbf{a})$ with $\mathbf{a}$ as above. Obviously, we have

$$1 \leqslant p(n, \xi, b) \leqslant b^n,$$

and both inequalities are sharp (take e.g. the analogue in base b of the number given in (1) to show that the right-hand inequality is sharp). A weaker conjecture than Conjecture 1 is the following.

Conjecture 2. *For every real irrational algebraic number ξ, every positive integer n and every base b, we have $p(n, \xi, b) = b^n$.*

It may also be stated under the following equivalent form.

Conjecture 2. *If there exist a positive integer n and a base b for which the irrational number ξ satisfies $p(n, \xi, b) < b^n$, then ξ is transcendental.*

It is easily seen that, if $\mathbf{a}$ is ultimately periodic, then the numbers $p(n, \xi, b)$ are uniformly bounded by a constant depending only on b. If this is not the case, then, by a result of Morse & Hedlund [12], [13], we have $p(n, \xi, b) \geqslant n+1$ for every positive integer n. Sequences $\mathbf{w}$ such that $p(n, \mathbf{w}) = n+1$ for every positive n do exist and are called *Sturmian sequences.* A first step towards a proof of Conjecture 2 would be to establish a good lower bound for $p(n, \xi, b)$ when ξ is irrational algebraic. The first result of this type was proved in 1997 by Ferenczi & Mauduit [8].

THEOREM FM. *If ξ is algebraic irrational, then the tail of the expansion of ξ in base b cannot be a Sturmian sequence.*

The Sturmian sequences can be viewed as the "simplest" non-periodic sequences. Thus, Theorem FM asserts that the b-adic expansion of every algebraic number cannot be "too simple".

Actually, the approach of [8] yields a slightly stronger result, namely that

$$\lim_{n \to \infty} \big(p(n, \xi, b) - n \big) = +\infty, \tag{2}$$

for any algebraic irrational number ξ (see [3]).

As explained in Section 2, Theorem FM follows from a specific combinatorial transcendence criterion. Recently, a new transcendence criterion was obtained by Adamczewski, Bugeaud & Luca [2], that yields (see [1]) a strong improvement of Theorem FM and of (2).

THEOREM 1. *If ξ is algebraic irrational, then*

$$\lim_{n \to \infty} \frac{p(n, \xi, b)}{n} = +\infty.$$

Theorem 1 is a small step towards a confirmation of Conjecture 2. Nevertheless, we are still very far away from what is expected, and we remain unable to confirm the following conjecture.

CONJECTURE 3. *For every real irrational algebraic number ξ and every base b with $b \geqslant 3$, we have $p(1, \xi, b) \geqslant 3$.*

In particular, we do not know whether there are algebraic irrational numbers on the usual tryadic Cantor set. This particular question goes back to Mahler [11].

2. Some remarks on the proofs

The common strategy for the proofs of Theorems FM and 1 is to start from a transcendence criterion and to show that, when the complexity function $p(n, \xi, b)$ is small enough, then ξ must be either rational, or transcendental. Maybe the most known transcendence criterion is the Roth theorem [15], asserting that any irrational number that is too well approximable by rational numbers must be transcendental, more precisely:

Theorem R. *Let θ be an irrational real number. If there exist a positive real number ε and infinitely many rational numbers $\frac{p}{q}$ such that*

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{q^{2+\varepsilon}},$$

then θ is transcendental.

Let us see how to apply Theorem R to prove that a specific real number with low complexity is transcendental. Consider the morphism σ defined on the set of words on the alphabet $\{0, 1\}$ by $\sigma(0) = 01$ and $\sigma(1) = 0$. Then, we have $\sigma^2(0) = 010, \sigma^3(0) = 01001$, $\sigma^4(0) = 01001010$, and the sequence $(\sigma^n(0))_{n \geqslant 0}$ converges to the infinite word

$$\mathbf{f} = 0100101001001010010 10\ldots, \tag{3}$$

usually called the *Fibonacci word*. For any $j \geqslant 1$, denote by f_j the j-th digit of $\mathbf{f}$. Let ξ be the real number whose 2-adic expansion is given by $\mathbf{f}$. Let also $(F_j)_{j \geqslant 0}$ denote the Fibonacci sequence, that is, the sequence starting with $F_0 = 0$, $F_1 = 1$, and satisfying $F_{j+2} = F_{j+1} + F_j$, for any $j \geqslant 0$. We observe from (3) that

$$\mathbf{f} = f_1 f_2 f_3 f_1 f_2 f_3 f_7 \ldots = f_1 f_2 f_3 f_4 f_5 f_1 f_2 f_3 f_4 f_5 f_1 f_{12} \ldots\ .$$

Actually, it is possible (and not too difficult) to prove that, for any integer $j \geqslant 4$, we have

$$\mathbf{f} = f_1 f_2 \ldots f_{F_j} f_1 f_2 \ldots f_{F_j} f_1 f_2 \ldots f_{F_{j-1}-2} f_{F_{j+2}-1} \ldots\ .$$

This shows that the rational number

$$\begin{aligned} r_j &= \frac{f_1}{2} + \frac{f_2}{2^2} + \ldots + \frac{f_{F_j}}{2^{F_j}} + \frac{f_1}{2^{F_j+1}} + \frac{f_2}{2^{F_j+2}} + \ldots + \frac{f_{F_j}}{2^{2F_j}} + \ldots \\ &= \left(\frac{f_1}{2} + \frac{f_2}{2^2} + \ldots + \frac{f_{F_j}}{2^{F_j}}\right) \frac{2^{F_j}}{2^{F_j} - 1} \end{aligned}$$

$$=: \frac{p_j}{2^{F_j}-1}$$

is a very good approximation to

$$\xi = \sum_{j\geqslant 1} \frac{f_j}{2^j}.$$

Indeed, since the first $F_{j+2}-2$ digits in the 2-adic expansions of ξ and of r_j are the same, we have

$$0 < |\xi - r_j| < \sum_{h\geqslant -1} 2^{-F_{j+2}-h} = 2^{-F_{j+2}+2}. \tag{4}$$

An easy induction shows that $F_{j+2} \geqslant 1.5\,F_{j+1}$ for any $j \geqslant 2$. Consequently, we infer from (4) that

$$\left|\xi - \frac{p_j}{2^{F_j}-1}\right| < \frac{4}{(2^{F_j}-1)^{2.25}},$$

and it follows from Theorem R that ξ is transcendental.

The key observation for this proof is that the infinite word **f** begins in infinitely many "more-than-squares". Let us now formalize what has been done above. The length of a finite word U, that is, the number of letters composing U, is denoted by $|U|$. For any positive integer ℓ, we write U^ℓ for the word $U\ldots U$ (ℓ times repeated concatenation of the word U). More generally, for any positive real number x, we denote by U^x the word $U^{[x]}U'$, where U' is the prefix of U of length $\lceil (x-[x])|U|\rceil$. Here, and in all what follows, $[y]$ and $\lceil y\rceil$ denote, respectively, the integer part and the upper integer part of the real number y. Let b, ξ and $\mathbf{a}=(a_n)_{n\geqslant 1}$ be as in Section 1. Let $w>1$ and $c\geqslant 0$ be real numbers. We say that **a** satisfies Condition $(*)_{c,w}$ if **a** is not eventually periodic and if there exist two sequences of finite words $(U_j)_{j\geqslant 1}$, $(V_j)_{j\geqslant 1}$ such that:

(i) For any $j\geqslant 1$, the word $U_jV_j^w$ is a prefix of the word **a**;
(ii) The sequence $(|U_j|/|V_j|)_{j\geqslant 1}$ is bounded from above by c;
(iii) The sequence $(|V_j|)_{j\geqslant 1}$ is strictly increasing.

We proved above that the Fibonacci word **f** satisfies Condition $(*)_{2.25,0}$. Furthermore, we have observed the following fact:

If the sequence **a** *satisfies Condition* $(*)_{w,0}$ *for some* $w>2$, *then the corresponding real number is transcendental.*

The idea of Ferenczi & Mauduit was to apply an extension of Roth's theorem to get a stronger transcendence criterion. Indeed, a disadvantage of the use of Theorem R is that we need that the repetitions occur at the very beginning (or we have to assume that w is much larger than 2). They showed that the p-adic extension of Roth's theorem due to Ridout [14] yields the following criterion:

If the sequence $\mathbf{a}$ *satisfies Condition* $(*)_{w,c}$ *for some* $w > 2$ *and some* $c \geqslant 0$*, then the associated real number is transcendental.*

They were able to prove that this criterion applies to every Sturmian sequence, which yields Theorem FM. It is still needed that "more-than-squares" occur, but they do not need anymore to occur at the very beginning, but still "not too far" from the beginning. Let us give some more explanation.

It is not difficult to see that, if $\mathbf{a}$ satisfies Condition $(*)_{w,c}$ for some $w > 2$ and some $c \geqslant 0$, then, for any $j \geqslant 1$, there exist integers p_j, r_j and s_j such that

$$\left| \xi - \frac{p_j}{b^{r_j}(b^{s_j} - 1)} \right| < \frac{2}{b^{r_j + ws_j}},$$

where $r_j = |U_j|$ and $s_j = |V_j|$. We see that the rational approximations of ξ obtained in that way are very specific: their denominators have a possibly very large part composed of fixed prime numbers (namely, the prime divisors of b). This is precisely what is exploited by Ferenczi & Mauduit. To prove that ξ is transcendental by applying Theorem R, we clearly need to check that

$$\limsup_{j \to +\infty} \frac{r_j + ws_j}{r_j + s_j} > 2,$$

or, equivalently, that

$$\limsup_{j \to +\infty} \frac{(w-2)s_j - r_j}{r_j + s_j} > 0. \tag{5}$$

The use of the Ridout Theorem [14] instead of Theorem R enabled Ferenczi & Mauduit to replace (5) by the weaker inequality

$$\limsup_{j \to +\infty} \frac{(w-2)s_j}{r_j + s_j} > 0. \tag{6}$$

The next step has been made by Adamczewski, Bugeaud & Luca [2], who used a p-adic version of the Schmidt Subspace Theorem (which is a multidimensional extension of Roth's Theorem, see [16], [17]) to get the following improved criterion.

If the sequence $\mathbf{a}$ *satisfies Condition* $(*)_{w,c}$ *for some* $w > 1$ *and some* $c \geqslant 0$*, then the associated real number is transcendental.*

They took advantage of the specific shape of the factors $b^{s_j} - 1$ of the denominators of the good approximations to ξ and managed to replace (6) by the weaker inequality

$$\limsup_{j \to +\infty} \frac{(w-1)s_j}{r_j + s_j} > 0. \tag{7}$$

It is easily seen that the above written criterion is a rephrasement of (7). In this new criterion, it is not needed anymore that squares occur in order to prove the transcendence of our number. Only slight stammerings are enough, provided that they do not occur too far from the beginning. As shown in [1], simple combinatoric arguments allow us to deduce Theorem 1 from this third criterion.

3. Discussion

A first important remark is that we cannot hope to prove Conjecture 2 with this method. Indeed, the above quoted theorems of Roth, Ridout and Schmidt show that algebraic numbers have a property shared by almost all numbers. Here, to show that our given number is transcendental, we simply prove that it does not share this property. In particular, the set of numbers which can be proved to be transcendental by using one of our criteria has Lebesgue measure zero.

When investigating these problems, it seems to be natural to introduce another notion of complexity, namely that of *asymptotic complexity.* Keeping the notation of Section 1, for any positive integer n, we let $\widehat{p}(n, \xi, b)$ denote the number of distinct blocks of n letters occurring infinitely often in $\mathbf{a}$. Obviously, we have

$$1 \leqslant \widehat{p}(n, \xi, b) \leqslant b^n,$$

and both inequalities are sharp. If ξ is irrational, then $\widehat{p}(n, \xi, b) \geqslant n + 1$ holds, by a result of Morse & Hedlund [12], [13]. A weaker conjecture than Conjecture 1 is the following.

Conjecture 4. *For every real irrational algebraic number ξ, every positive integer n and every base b, we have $\widehat{p}(n, \xi, b) = b^n$.*

Even a much weaker conjecture remains unanswered.

Conjecture 5. *For every real irrational algebraic number ξ, every positive integer n and every base b, we have $\widehat{p}(n, \xi, b) > n + 1$.*

If proved, then Conjecture 5 would imply that the (irrational) number ξ whose 2-adic expansion is given by

$$010010001000010000010000000100\ldots$$

is transcendental. It is known that ξ is not a quadratic real number (see [9] for a proof and [4] for related results). Actually, ξ is transcendental: this is a consequence of deep transcendence results proved by D. Duverney *et al.*, concerning the values of theta series at algebraic points [7]. Notice that there are positive constants c_1 and c_2 such that $c_1 n^2 \leqslant p(n, \xi, 2) \leqslant c_2 n^2$, for any $n \geqslant 1$: the complexity is too large and we cannot apply a transcendence criterion given in Section 2. Notice also that $\widehat{p}(n, \xi, 2) = n + 1$ for any $n \geqslant 1$.

To conclude, we quote a result of Mahler [10] related, in some way, to Conjecture 4.

Theorem M. *For every base b and any positive integer n, there exist algebraic irrational numbers ξ such that $\widehat{p}(n, \xi, b) = b^n$.*

Actually, Theorem M is a consequence of a general result on the digits of the integer multiples of a given irrational number, see also [5].

References

[1] B. Adamczewski, Y. Bugeaud, On the complexity of algebraic numbers I. Expansions in integer bases, *Ann. of Math.* (to appear).

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

[3] J.-P. Allouche, Nouveaux résultats de transcendance de réels à développements non aléatoire, *Gaz. Math.*, **84**, 19–34 (2000).

[4] D. H. Bailey, J. M. Borwein, R. E. Crandall, C. Pomerance, On the binary expansions of algebraic numbers, *J. Théor. Nombres Bordeaux*, **16**, 487–518 (2004).

[5] D. Berend, M. D. Boshernitzan, On a result of Mahler on the decimal expansions of $(n\alpha)$, *Acta Arith.*, **66**, 315–322 (1994).

[6] D. G. Champernowne, The construction of decimals normal in the scale of ten, *J. London Math. Soc.*, **8**, 254–260 (1933).

[7] D. Duverney et al., Transcendence of Jacobi's theta series, *Proc. Japan Acad. Ser. A*, **72**, 202–203 (1996).

[8] S. Ferenczi, C. Mauduit, Transcendence of numbers with a low complexity expansion, *J. Number Theory*, **67**, 146–161 (1997).

[9] F. Luca, .121221222 . . . is not quadratic, *Rev. Mat. Complut.*, **18**, 353–362 (2005).

[10] K. Mahler, Arithmetical properties of the digits of the multiples of an irrational number, *Bull. Austral. Math. Soc.*, **8**, 191–203 (1973).

[11] K. Mahler, Some suggestions for further research, *Bull. Austral. Math. Soc.*, **29**, 101–108 (1984).

[12] M. Morse, G. A. Hedlund, Symbolic dynamics, *Amer. J. Math.*, **60**, 815–866 (1938).

[13] M. Morse, G. A. Hedlund, Symbolic dynamics. II, *Amer. J. Math.*, **62**, 1–42 (1940).

[14] D. Ridout, Rational approximations to algebraic numbers, *Mathematika*, **4**, 125–131 (1957).

[15] K. F. Roth, Rational approximations to algebraic numbers, *Mathematika*, **2**, 1–20 (1955); corrigendum, 168.

[16] H. P. Schlickewei, On products of special linear forms with algebraic coefficients, *Acta Arith.*, **31**, 389–398 (1976).

[17] W. M. Schmidt, Diophantine approximation, *Lecture Notes in Math.*, **785**, Springer (1980).

Received
30/5/2005