On the independence of expansions of algebraic numbers in an integer base

Boris Adamczewski (Lyon) & Yann Bugeaud (Strasbourg)

Abstract. *Let $b \geq 2$ be an integer. According to a conjecture of Émile Borel, the b-adic expansion of any irrational algebraic number behaves in some respect 'like a random sequence'. We give a contribution to the following related problem: let α and α' be irrational algebraic numbers, prove that their b-adic expansions either have the same tail, or behave in some respect 'like independent random sequences'.*

1. Introduction

Let $b \geq 2$ be an integer. Despite some recent progress [1,2,3], the behaviour of the b-adic expansion of every irrational algebraic number α is still rather mysterious. According to Émile Borel [4], such an expansion should satisfy some of the same laws as do almost all real numbers. In particular, it is expected that every algebraic real number α is normal in base b. This means that, for any positive integer n, each one of the b^n blocks of length n on the alphabet $\{0, 1, \ldots, b-1\}$ should occur in the b-adic expansion of α with the same frequency $1/b^n$. Thus, a particular case of Émile Borel's conjecture essentially claims that the b-adic expansions of $\sqrt{2}$ and $\sqrt{3}$ behave in some respect 'like random sequences'. In view of this, we can go a step further and ask whether these expansions behave, in some sense, like two *independent* random sequences.

Let make this idea more precise. Given a real number $C < \sqrt{b}$, almost every pair (α, α') of real numbers in $(0, 1)$ is independent in the following sense: there are only finitely many indices n such that the prefixes of length $[C^n]$ of the b-adic expansions of α and α' have a common block of length n. This easily follows from the Cantelli lemma, as shown in Section 5 below. Motivated by Émile Borel's conjecture, we ask whether every pair (α, α') of algebraic numbers, not both rational, is independent in the above sense. Observe first that the answer is trivially negative when the b-adic expansions of α and α' have the same tail. However, this case excepted, we feel that the answer should probably be positive. The main purpose of the present work is to prove a first result towards a confirmation of this guess (see Theorem 1).

Say differently, we study how close to each other the b-adic expansions of two distinct irrational algebraic numbers can be. In particular, our results show that if one slightly

2000 *Mathematics Subject Classification : 11A63, 11K16, 11J68*

perturbs the b-adic expansion of an algebraic number α and gets a number α', then α' is transcendental, except in the trivial case when the tails of the expansions of α and α' coincide (see Theorem 2). As a consequence, we derive a new proof of the combinatorial transcendence criterion given in [2]. Note that this transcendence criterion was used in [1] to prove, among other results, the so-called Cobham–Loxton–van der Poorten conjecture, claiming that the b-adic expansion of every algebraic irrational number cannot be generated by a finite automaton. Our proof rests on a p-adic version of the Schmidt Subspace Theorem established by Schlickewei [6].

2. Main result

Throughout the paper, b denotes an integer at least equal to 2, and all the algebraic numbers we consider lie in the interval $(0, 1)$. This does not restrict the generality.

For any α in $(0, 1)$ there is a unique sequence $(a_k)_{k\geq 1}$ of integers from $\{0, 1, \ldots, b-1\}$ such that $\alpha = \sum_{k\geq 1} a_k/b^k$ and a_k is non-zero for infinitely many positive integers k. We call this sequence the b-adic expansion of α. Sometimes, it is convenient to view it as the infinite word $\mathbf{a} = a_1 a_2 \ldots$ on the alphabet $\{0, 1, \ldots, b-1\}$. In particular, for any integer $k \geq 1$, the prefix of length k of the b-adic expansion of α is the word $a_1 \ldots a_k$.

We begin with three definitions.

Definition 1. *Let $f : \mathbf{Z}_{\geq 1} \to \mathbf{Z}_{\geq 1}$ be any increasing function. The pair (α, α') of real numbers is called f-dependent if there exist infinitely many integers n for which the prefixes of length $f(n)$ of their b-adic expansions have in common a block of length n. Otherwise, it is called f-independent.*

Note that if (α, α') denotes an f-independent pair of real numbers, and if g is an increasing function such that $f(n) \geq g(n)$ for every positive integer n large enough, then (α, α') is also g-independent.

Definition 2. *Two real numbers α and α' are said to be equivalent if their b-adic expansions have the same tail.*

Definition 3. *Let $f : \mathbf{Z}_{\geq 1} \to \mathbf{Z}_{\geq 1}$ be an increasing function. We say that f is a function with linear growth if there exists a positive constant c such that $f(n) < cn$, for every integer n.*

Our main result can then be stated as follows.

Theorem 1. *Let $f : \mathbf{Z}_{\geq 1} \to \mathbf{Z}_{\geq 1}$ be any increasing function with linear growth. Let α and α' be two algebraic numbers in $(0, 1)$. Then, either α and α' are equivalent, or the pair (α, α') is f-independent.*

Most probably, the conclusion of Theorem 1 should remain true when f has polynomial growth, and even when f increases more slowly than any exponential function.

3. A useful transcendence criterion

In this Section, we present a transcendence criterion from which we will derive our main result, namely Theorem 1.

We first introduce some notation. It is convenient to use the terminology from combinatorics on words. Let $\mathcal{A}$ be a finite set. The length of a word W on the alphabet $\mathcal{A}$, that is, the number of letters composing W, is denoted by $|W|$. For any positive integer ℓ, we write W^ℓ for the word $W \dots W$ (ℓ times repeated concatenation of the word W). More generally, for any positive real number x, we denote by W^x the word $W^{[x]}W'$, where W' is the prefix of W of length $\lceil (x-[x])|W| \rceil$. 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 $\mathbf{a} = (a_k)_{k\geq 1}$ be a sequence of elements from $\mathcal{A}$, that we identify with the infinite word $a_1 a_2 \dots$

We say that $\mathbf{a}$ is a *stammering sequence* if there exist a real number $w > 1$ and two sequences of finite words $(W_n)_{n\geq 1}$, $(X_n)_{n\geq 1}$ such that:

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

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

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

The main result of [2] then reads as follows.

Theorem ABL. *Let $\mathbf{a} = (a_k)_{k\geq 1}$ be a stammering sequence of integers from $\{0, 1, \dots, b-1\}$. Then, the real number*

$$\alpha = \sum_{k=1}^{+\infty} \frac{a_k}{b^k}$$

is either rational or transcendental.

Let $\mathbf{a} = (a_k)_{k\geq 1}$ and $\mathbf{a}' = (a'_k)_{k\geq 1}$ be sequences of elements from $\mathcal{A}$, that we identify with the infinite words $a_1 a_2 \dots$ and $a'_1 a'_2 \dots$, respectively. We say that the pair $(\mathbf{a}, \mathbf{a}')$ satisfies Condition $(*)$ if there exist three sequences of finite words $(U_n)_{n\geq 1}$, $(U'_n)_{n\geq 1}$, and $(V_n)_{n\geq 1}$ such that:

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

(ii) For any $n \geq 1$, the word $U'_n V_n$ is a prefix of the word $\mathbf{a}'$;

(iii) The sequences $(|U_n|/|V_n|)_{n\geq 1}$ and $(|U'_n|/|V_n|)_{n\geq 1}$ are bounded from above;

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

If, moreover, we add the condition

(v) The sequence $(|U_n| - |U'_n|)_{n\geq 1}$ is unbounded,

then, we say that the pair $(\mathbf{a}, \mathbf{a}')$ satisfies Condition $(**)$.

Theorem 2. *Let $\mathbf{a} = (a_k)_{k\geq 1}$ and $\mathbf{a}' = (a'_k)_{k\geq 1}$ be sequences of integers from $\{0, 1, \ldots, b-1\}$. If the pair $(\mathbf{a}, \mathbf{a}')$ satisfies Condition $(*)$, then at least one of the real numbers*

$$\alpha := \sum_{k=1}^{+\infty} \frac{a_k}{b^k}, \quad \alpha' := \sum_{k=1}^{+\infty} \frac{a'_k}{b^k}$$

is transcendental, or α and α' are equivalent. Furthermore, if the pair $(\mathbf{a}, \mathbf{a}')$ satisfies Condition $(**)$, then at least one of the real numbers α, α' is transcendental, or they are equivalent and both rational.

Theorem 2 implies both Theorem 1 and Theorem ABL, as shown in the next Section. In particular, the present approach provides a new proof of Theorem ABL.

4. Proofs of Theorems 1, 2 and ABL

The proofs of our results rest on the following p-adic version [6] of the Schmidt Subspace Theorem [7]. Throughout the present note, the p-adic absolute value $|\cdot|_p$ is normalized such that $|p|_p = p^{-1}$.

Theorem A. *Let $m \geq 2$ be an integer. Let S be a finite set of places on $\mathbf{Q}$ containing the infinite place. Let $L_{1,\infty}, \ldots, L_{m,\infty}$ be m linearly independent linear forms with real algebraic coefficients. For any finite place v in S, let $L_{1,v}, \ldots, L_{m,v}$ be m linearly independent linear forms with integer coefficients. Let ε be a positive real number. Then, the set of solutions $\mathbf{x} = (x_1, \ldots, x_m)$ in $\mathbf{Z}^m$ to the inequality*

$$\prod_{v\in S} \prod_{i=1}^{m} |L_{i,v}(\mathbf{x})|_v \leq (\max\{[x_1|, \ldots, |x_m|\})^{-\varepsilon}$$

lies in finitely many proper subspaces of $\mathbf{Q}^m$.

Proof : This is a particular case of Theorem 4.1 from [6]. □

We first establish Theorem 2.

Proof of Theorem 2. We keep the notation of this theorem. We assume that α and α' are both algebraic numbers and we aim at proving that if the pair $(\mathbf{a}, \mathbf{a}')$ satifies Condition $(*)$ (resp. Condition $(**)$), then α and α' are equivalent (resp. equivalent and both rational).

We further set $|U_n| = r_n$, $|U'_n| = r'_n$ and $|V_n| = s_n$, for $n \geq 1$. Let d_n and d'_n be the rational integers defined by

$$\sum_{k=1}^{r_n+s_n} \frac{a_k}{b^k} = \frac{d_n}{b^{r_n+s_n}} \quad \text{and} \quad \sum_{k=1}^{r'_n+s_n} \frac{a'_k}{b^k} = \frac{d'_n}{b^{r'_n+s_n}}.$$

The key observation is that, for any prime number p dividing b, the p-adic distance between d_n and d'_n is very small. Indeed, we have

$$|d_n - d'_n|_p \leq |b|_p^{s_n}.$$

Consider now the four linearly independent linear forms with real algebraic coefficients:

$$\begin{aligned} L_{1,\infty}(X_1, X_2, X_3, X_4) =& \alpha X_1 - X_3, \\ L_{2,\infty}(X_1, X_2, X_3, X_4) =& \alpha' X_2 - X_4, \\ L_{3,\infty}(X_1, X_2, X_3, X_4) =& X_1, \\ L_{4,\infty}(X_1, X_2, X_3, X_4) =& X_2. \end{aligned}$$

Evaluating them on the integer points $\mathbf{x}_n := (b^{r_n+s_n}, b^{r'_n+s_n}, d_n, d'_n)$, we get that

$$\prod_{1\le j\le 4} |L_{j,\infty}(\mathbf{x}_n)| \le b^{r_n+r'_n+2s_n}. \tag{1}$$

For any prime number p dividing b, we consider the four linearly independent linear forms with integer coefficients:

$$\begin{aligned} L_{1,p}(X_1, X_2, X_3, X_4) =& X_1, \\ L_{2,p}(X_1, X_2, X_3, X_4) =& X_2, \\ L_{3,p}(X_1, X_2, X_3, X_4) =& X_3, \\ L_{4,p}(X_1, X_2, X_3, X_4) =& X_4 - X_3. \end{aligned}$$

We get that

$$\prod_{p|b} \prod_{1\le j\le 4} |L_{j,p}(\mathbf{x}_n)|_p \le b^{-(r_n+r'_n+2s_n)}\, b^{-s_n}. \tag{2}$$

It follows from (1), (2), and assumption (iii) from the definition of Condition $(*)$ that there exists a positive real number ε such that

$$\begin{aligned} \prod_{1\le j\le 4} |L_{j,\infty}(\mathbf{x}_n)| \cdot \prod_{p|b} \prod_{1\le j\le 4} |L_{j,p}(\mathbf{x}_n)|_p &\le b^{-s_n} \\ &\le \max\{b^{r_n+s_n}, b^{r'_n+s_n}, d_n, d'_n\}^{-\varepsilon}, \end{aligned}$$

for every $n \ge 1$.

We then infer from Theorem A that all the points $\mathbf{x}_n$ lie in a finite number of proper subspaces of $\mathbf{Q}^4$. Thus, there exist a non-zero integer quadruple (z_1, z_2, z_3, z_4) and an infinite set of distinct positive integers $\mathcal{N}_1$ such that

$$z_1 b^{r_n+s_n} + z_2 b^{r'_n+s_n} + z_3 d_n + z_4 d'_n = 0, \tag{3}$$

for any n in $\mathcal{N}_1$.

Assume first that $(r_n - r'_n)_{n\in\mathcal{N}_1}$ is bounded. Then, there exist an integer r and infinitely many integers n in $\mathcal{N}_1$ for which $r_n - r'_n = r$. Without any loss of generality, we may assume that $r \ge 0$. Consequently, there is a finite (possibly empty) word W of length r, a sequence of finite words $(W_n)_{n\ge 1}$, and an infinite set of distinct positive integers $\mathcal{N}_2$ such that, for any n in $\mathcal{N}_2$, the word $\mathbf{a}$ begins in WW_nV_n, the word $\mathbf{a}'$ begins in U'_nV_n, and

$|W_n| = |U'_n| = r'_n$. This implies that there exist an integer ℓ and a sequence of integers $(\ell'_n)_{n\geq 1}$ such that

$$|b^{r'_n}\alpha' - b^{r'_n}(b^r\alpha - \ell) - \ell'_n| \leq b^{-s_n}, \quad \text{for } n \text{ in } \mathcal{N}_2,$$

and, by assumption (iii) from the definition of Condition $(*)$, there exists a positive real number ε such that

$$A_n := \left|\alpha' - (b^r\alpha - \ell) - \frac{\ell'_n}{b^{r'_n}}\right| \leq \frac{1}{b^{(1+\varepsilon)r'_n}}, \quad \text{for } n \text{ in } \mathcal{N}_2.$$

We then infer from Ridout's theorem [5] that $A_n = 0$ for n large enough in $\mathcal{N}_2$. Indeed, otherwise $\alpha' - (b^r\alpha - \ell)$ would be transcendental which is in contradiction with the assumption that α and α' are both algebraic. Now, it follows from $A_n = 0$ (for some given integer n) that the sequences $\mathbf{a}$ and $\mathbf{a}'$ have the same tail.

Assume now that $(r_n - r'_n)_{n\in\mathcal{N}_1}$ is unbounded. Without any loss of generality, we may suppose that $r_n - r'_n$ tends to infinity with n. Dividing (3) by $b^{r_n+s_n}$, we get

$$z_1 + z_2 b^{r'_n - r_n} + z_3 \frac{d_n}{b^{r_n+s_n}} + z_4 b^{r'_n - r_n} \frac{d'_n}{b^{r'_n+s_n}} = 0. \tag{4}$$

Letting n tend to infinity along $\mathcal{N}_1$, we infer from (4) that either α is rational, or $z_1 = z_3 = 0$. In the latter case, we obtain that α' is rational. Let us assume that α is rational, the case when α' is rational being similar.

Then, the sequence $\mathbf{a}$ is eventually periodic. There exist a finite (possibly empty) word U of length m and an infinite set of distinct positive integers $\mathcal{N}_3$ such that, for any n in $\mathcal{N}_3$, the word $\mathbf{a}$ begins in UV_n while the word $\mathbf{a}'$ begins in U'_nV_n. This implies that there exist a rational number p/q and a sequence of integers $(\ell'_n)_{n\geq 1}$ such that

$$\left|b^{r'_n}\alpha' - \frac{p}{q} - \ell'_n\right| \leq b^{-s_n}, \quad \text{for } n \text{ in } \mathcal{N}_3.$$

As above, we infer from Ridout's theorem that $q\alpha'$ is then rational, hence, that α' is rational. Consequently, the sequences $\mathbf{a}$ and $\mathbf{a}'$ are both eventually periodic. Moreover, since by (i), (ii) and (iv) these two sequences have arbitrarily large common factors, they have the same tail. This achieves the proof. □

Proof of Theorem 1. Let $f : \mathbf{Z}_{\geq 1} \to \mathbf{Z}_{\geq 1}$ be an increasing function with linear growth. Assume that α and α' are non-equivalent real numbers, and that the pair (α, α') is f-dependent. We have to prove that at least one of the numbers α, α' is transcendental. Let denote by $\mathbf{a}$ (resp. $\mathbf{a}'$) the b-adic expansion of α (resp. α'). In view of Theorem 2, it is sufficient to show that the pair $(\mathbf{a}, \mathbf{a}')$ satisfies Condition $(*)$.

By assumption, there exists an increasing sequence of positive integers $(n_k)_{k\geq 1}$ such that for every $k \geq 1$, the prefixes of length $f(n_k)$ of $\mathbf{a}$ and $\mathbf{a}'$ have a common block of length n_k, that we denote by V_k. In other words, there exist two sequences $(U_k)_{k\geq 1}$ and $(U'_k)_{k\geq 1}$ such that

(i) the word U_kV_k is a prefix of the word $\mathbf{a}$;

(ii) the word U'_kV_k is a prefix of the word $\mathbf{a}'$;

(ii') $|U_kV_k| \leq f(n_k)$ and $|U'_kV_k| \leq f(n_k)$.

We thus infer from (ii'), $|V_k| = n_k$, and the fact that f is a function with linear growth that (iii) from the definition of Condition $(*)$ is satisfied. Furthermore, condition (iv) is also satisfied since $(n_k)_{k\geq 1}$ is an increasing sequence. Consequently, the pair $(\mathbf{a}, \mathbf{a}')$ satisfies Condition $(*)$. This ends the proof. □

We end this Section with an alternative proof of Theorem ABL.

Proof of Theorem ABL. Let $\mathbf{a} = (a_k)_{k\geq 1}$ be a stammering sequence with values in $\{0, 1, \ldots, b-1\}$ and set $\alpha = \sum_{k=1}^{+\infty} a_k/b^k$. In view of Theorem 2, it is sufficient to prove that the pair $(\mathbf{a}, \mathbf{a})$ satisfies Condition $(**)$.

We assume that the real number w, and the sequences $(W_n)_{n\geq 1}$ and $(X_n)_{n\geq 1}$ occurring in the definition of a stammering sequence are fixed. Let $n \geq 1$ be an integer.

If $w \geq 2$, we set $U_n = W_n$, $V_n = X_n$ and $U'_n = W_nX_n$, and we easily get that the pair $(\mathbf{a}, \mathbf{a})$ satisfies Condition $(**)$.

If $1 < w < 2$, then there exists a unique prefix X'_n of X_n such that $X_n^w = X_nX'_n$. Then, we set $U_n = W_n$, $V_n = X'_n$ and $U'_n = W_nX_n$, and we easily get that the pair $(\mathbf{a}, \mathbf{a})$ satisfies Condition $(**)$. This ends the proof. □

5. A metrical appendix

In the present Section, we justify a claim from the introduction.

Proposition 1. *Let $b \geq 2$ be an integer. For any real number $C < \sqrt{b}$, almost every pair (α, α') of real numbers in $(0, 1)$ is f-independent, where the function f is defined by $f(n) = [C^n]$ for every positive integer n.*

Proof : Let n be a positive integer and let $C > 1$ be as above. The measure of the set of real pairs (α, α') having a common block of length n in the prefixes of length $[C^n]$ of their b-adic expansions is at most equal to $C^{2n} \cdot b^{-n}$. Our assumption implies that the series $\sum_{n\geq 1}(b/C^2)^{-n}$ converges, thus, by the Cantelli lemma, almost no pair (α, α') of real numbers in $(0, 1)$ is f-dependent. This proves the proposition. □

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** (2004), 11–14.

[3] D. H. Bailey, J. M. Borwein, R. E. Crandall & C. Pomerance, On the binary expansions of algebraic numbers, *J. Théor. Nombres Bordeaux* **16** (2004), 487–518.
[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] D. Ridout, Rational approximations to algebraic numbers, *Mathematika* **4** (1957), 125–131.
[6] H. P. Schlickewei, On products of special linear forms with algebraic coefficients, *Acta Arith.* **31** (1976), 389–398.
[7] W. M. Schmidt, *Diophantine approximation*, Lecture Notes in Mathematics **785**, Springer, 1980.

Boris Adamczewski
CNRS, Institut Camille Jordan
Université Claude Bernard Lyon 1
Bât. Braconnier, 21 avenue Claude Bernard
69622 VILLEURBANNE Cedex (FRANCE)
Boris.Adamczewski@math.univ-lyon1.fr

Yann Bugeaud
Université Louis Pasteur
U. F. R. de mathématiques
7, rue René Descartes
67084 STRASBOURG Cedex (FRANCE)
bugeaud@math.u-strasbg.fr