An explicit lower bound for the block complexity of an algebraic number

Yann Bugeaud (Strasbourg)

Abstract. *Let $b \ge 2$ be an integer and ξ be an irrational real number. Among other results, we establish an explicit lower bound for the number of distinct blocks of n digits occurring in the b-ary expansion of ξ.*

1. Introduction

Throughout the present note, b always denotes an integer at least equal to 2. Let ξ be a real number. There exist a rational integer A and a unique infinite sequence $\mathbf{a} = (a_j)_{j \ge 1}$ of integers from $\{0, 1, \ldots, b-1\}$ such that

$$\xi = A + \sum_{j \ge 1} \frac{a_j}{b^j}$$

and $\mathbf{a}$ does not terminate in an infinite string of digits $b-1$. Clearly, the sequence $\mathbf{a}$ is ultimately periodic if, and only if, ξ is rational. With a slight abuse of notation, we also denote by $\mathbf{a}$ the infinite word $a_1 a_2 \ldots$ and call it the b-ary expansion of ξ. 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 $\mathbf{a}$. To this end, for a finite or infinite word $\mathbf{w}$ on the alphabet $\{0, 1, \ldots, b-1\}$ and for a 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

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

and both inequalities are sharp.

Assume now that ξ is irrational and algebraic. Ferenczi and Mauduit [9] (see also [5]) proved in 1997 that its complexity function $n \mapsto p(n, \xi, b)$ satisfies

$$\lim_{n \to +\infty} \bigl(p(n, \xi, b) - n \bigr) = +\infty. \tag{1.1}$$

This result was recently considerably improved in [3], where

$$\lim_{n \to +\infty} \frac{p(n, \xi, b)}{n} = +\infty \tag{1.2}$$

2000 *Mathematics Subject Classification : 11J68, 11A63* *Keywords:* Transcendence, Schmidt Subspace Theorem, combinatorics on words

is established. The proof of (1.1) rests on the Ridout Theorem (that is, on a p-adic extension of the Roth Theorem), while that of (1.2) uses a p-adic extension of the Schmidt Subspace Theorem [11], worked out by Schlickewei. Consequently, (1.1) and (1.2) are ineffective in the sense that, for a given positive c, their proofs do not yield explicit values for n_1 and n_2 such that $p(n, \xi, b) \geq n + c$ for $n \geq n_1$ and $p(n, \xi, b) \geq cn$ for $n \geq n_2$. The question to know whether (1.2) can be made effective was posed to me by Sergei Konyagin at the conference on uniform distribution held in Luminy in January 2008. The purpose of the present note is to establish an effective (and explicit) version of (1.1) and to discuss further related effective results.

2. Results

Our first result asserts that if a long prefix of a real algebraic irrational number has small complexity (this situation can obviously happen, since algebraic irrational numbers form a dense subset of the real numbers), then its height and its degree cannot be both very small. Throughout the present note, the height $H(\alpha)$ of an algebraic number α is the maximum of the absolute values of the coefficients of its minimal defining polynomial over the integers.

Theorem 1. *Let $b \geq 2$ be an integer. Let ξ be a real algebraic irrational number of degree d and height at most H with $H \geq e^e$. Denote by $\mathbf{a}$ its b-ary expansion viewed as an infinite word on $\{0, 1, \ldots, b-1\}$. Let $\mathbf{w}$ be an infinite word whose complexity function satisfies $p(n, \mathbf{w}) \leq Cn$ for some integer $C \geq 2$ and all $n \geq 1$. Assume that the first L digits of $\mathbf{a}$ coincide with the first L digits of $\mathbf{w}$. Then we have*

$$H \geq \exp\bigl\{10^{-2} C^{-1} L^{1/(8 \log(4C))}\bigr\} \tag{2.1}$$

or

$$d \geq \exp\{10^{-100} C^{-11/2} (\log C)^{-1} (\log L)^{1/2} (\log \log L)^{-1}\}. \tag{2.2}$$

Theorem 1 gives an effective (but not very efficient!) procedure to test whether some real numbers given by their b-ary expansion can be algebraic of small height and small degree.

Our next statement implies an explicit version of the result of Ferenczi and Mauduit.

Theorem 2. *Let $b \geq 2$ be an integer. Let ξ be a real algebraic irrational number of degree d and height at most H with $H \geq e^e$. Set*

$$M = \exp\{10^{190} (\log(8d))^2 (\log \log(8d))^2\} + 2^{32 \log(240 \log(4H))}.$$

Then we have

$$p(n, \xi, b) \geq \left(1 + \frac{1}{M}\right) n, \quad \textit{for } n \geq 1. \tag{2.3}$$

Unfortunately, the present methods do not seem to be powerful enough to get an effective version of (1.2).

No importance has to be attached to the numerical constants occurring in Theorems 1 and 2. They can be slightly reduced.

Roughly speaking, the Subspace Theorem asserts that the set of integer solutions $(x_1, \ldots, x_n)$ to some given system of inequalities lies in finitely many proper subspaces of $\mathbf{Q}^n$. It is ineffective in the sense that we do not have an effective upper bound for $\max\{|x_1|, \ldots, |x_n|\}$. However, the Quantitative Subspace Theorem [12, 7] does provide us with an explicit upper bound for the number of exceptional proper subspaces in which all these solutions are contained. To our knowledge, this is the only available tool to get effective results in our context, and we use it in the proofs of Theorems 1 and 2.

We stress that the lower bounds (2.1), (2.2) and (2.3) do not depend on the base b. This is a consequence of the use of the Parametric Subspace Theorem, as in [6].

The proof of Theorem 2 uses the fact that the function $n \mapsto p(n, \xi, b)$ is strictly increasing when ξ is irrational. Actually, not much can be said on the behaviour of this function. Indeed, Ferenczi [8] established the existence of an infinite word $\mathbf{w}$ over a finite alphabet whose complexity function satisfies

$$\liminf_{n \to \infty} \frac{p(n, \mathbf{w})}{n} = 2 \quad \text{and} \quad \limsup_{n \to \infty} \frac{p(n, \mathbf{w})}{n^t} = +\infty \quad \text{for any } t > 1.$$

We stress that not every increasing function satisfying some obvious necessary conditions can be the block-complexity function of some infinite sequence.

Recall that the infinite word $\mathbf{w}$ is called quasi-Sturmian if there exist positive integers k and n_0 such that $p(n, \mathbf{w}) = n + k$ for all integers $n \geq n_0$. Besides that collection of infinite words, there also exist infinite words $\mathbf{w}$ with the property that

$$p(n, \mathbf{w}) - n \to +\infty \quad \text{and} \quad \lim_{n \to +\infty} \frac{p(n, \mathbf{w})}{n} = 1,$$

as established by Aberkane [2]. Furthermore, Aberkane [1] proved that, for any real number δ with $\delta > 1$ there exists an infinite word $\mathbf{w}$ satisfying

$$1 < \liminf_{n \to +\infty} \frac{p(n, \mathbf{w})}{n} < \limsup_{n \to +\infty} \frac{p(n, \mathbf{w})}{n} \leq \delta.$$

Both results show that our Theorem 2 gives in fact a stronger result than just an explicit version of (1.1).

3. Proofs

We begin with an auxiliary, combinatorial lemma. Throughout this section, we denote by $\lfloor \cdot \rfloor$ the integer part function.

Lemma 1. *Let $b \geq 2$ be an integer. Let ξ be a real number with $0 < \xi < 1$ and denote by $\mathbf{a}$ its b-ary expansion viewed as an infinite word on $\{0, 1, \ldots, b-1\}$. Assume that there are positive integers m_0, m_1 and $c \geq 2$ such that, for every integer n with $m_0 \leq n < m_1$, there is a word of length n having two occurrences in the prefix of length cn of $\mathbf{a}$. Setting*

$$N = \lfloor \log(m_1/m_0)/\log(4c+4) \rfloor,$$

there exist non-negative integers $p_1, \ldots, p_N, r_1, \ldots, r_N$ and positive integers $s_1, \ldots, s_N$ such that $r_1 + s_1 \geq m_0/2$, $s_1 < \ldots < s_N$, and

(i) $r_j \leq (2c+1)s_j, \quad (j = 1, \ldots, N)$;

(ii) *If $r_j \geq 1$, then b does not divide p_j,* $\quad (j = 1, \ldots, N)$;

(iii) $2(r_j + s_j) \leq r_{j+1} + s_{j+1} \leq 8c^2(r_j + s_j), \quad (j = 1, \ldots, N-1)$;

(iv)
$$\left| \xi - \frac{p_j}{b^{r_j}(b^{s_j} - 1)} \right| < \frac{1}{b^{(1+1/(5c))(r_j+s_j)}}, \quad (j = 1, \ldots, N).$$

Proof. This follows from an easy modification of the proof of Proposition 10.1 from [4]. We omit the details. □

Now, we discuss consequences of Lemma 1 and the Parametric Subspace Theorem [7]. We content ourself to sketch the proofs of our theorems, since they are very similar to that of Theorem 2.1 from [6].

Let ξ be a real, algebraic irrational number of degree d with $0 < \xi < 1$. Let $H \geq e^e$ be an upper bound for its height. Let m_0, m_1 and c be as in the statement of Lemma 1, and set $N = \lfloor \log(m_1/m_0)/\log(4c+4) \rfloor$.

Preparation for the proofs of Theorems 1 and 2. We assume that the hypotheses of Lemma 1 are satisfied. It follows from (iv) that, for $n = \lfloor N/2 \rfloor, \ldots, N$, the vector

$$\mathbf{x}_n := (b^{t_n}, b^{r_n}, p_n),$$

where we have set $t_n = r_n + s_n$, satisfies a system of inequalities to which we can apply the Parametric Subspace Theorem with $\varepsilon = 1/(5c)$, exactly as in [6]. We use an explicit estimate for the number of subspaces that contain all the solutions having a sufficiently large height. For this reason, we need to consider only points with large height. Thus, we assume that

$$N \geq 3 \log\big(80c \log(4H)\big). \tag{3.1}$$

Arguing as in [6], we establish that all the vectors $\mathbf{x}_n$, $\lfloor N/2 \rfloor \leq n \leq N$, lie in the union of at most

$$10^{160} c^8 \log(8d) \log\log(8d)$$

proper rational linear subspaces of $\mathbf{Q}^3$.

Let then $z_1X_1+z_2X_2+z_3X_3=0$ be such a subspace $\mathcal{H}$, where (z_1,z_2,z_3) is a non-zero primitive triple of rational integers. Let

$$\mathcal{N}=\{i_1<i_2<\ldots<i_r\}$$

be the set of n with $\lfloor N/2\rfloor \le n \le N$ such that $\mathbf{x}_n$ is in $\mathcal{H}$. Arguing again as in [6] and using (3.1), it follows that

$$r \le 10^{20}c^3(\log c)\log(8d)\log\log(8d)$$

Consequently, we get the bound

$$N \le 10^{180}c^{11}(\log c)(\log(8d))^2(\log\log(8d))^2, \tag{3.2}$$

provided that (3.1) holds. Combining (3.1) and (3.2), we have proved that

$$\begin{aligned}\lfloor \log(m_1/m_0)/\log(4c+4)\rfloor \le \max\{&3\log\bigl(80c\log(4H)\bigr),\\ &10^{180}c^{11}(\log c)(\log(8d))^2(\log\log(8d))^2\}.\end{aligned} \tag{3.3}$$

We are now in position to establish our theorems.

Proof of Theorem 1. Assume that $m_0=1$. Then, we get that either

$$m_1 \le (4c+4)^{4\log(80c\log(4H))},$$

or

$$m_1 \le \exp\{10^{181}c^{11}(\log c)^2(\log(8d))^2(\log\log(8d))^2\}.$$

By Dirichlet's *Schubfachprinzip*, an infinite word whose complexity functon is bounded by Cn satisfies the assumption of Lemma 1 with $c=C+1$, $m_0=1$ and m_1 arbitrary. This gives Theorem 1. $\square$

Proof of Theorem 2. Set $c=3$ and

$$\begin{aligned}T &= \exp\{10^{180}c^{11}(\log c)^2(\log(8d))^2(\log\log(8d))^2\}+(4c+4)^{4\log(80c\log(4H))}\\ &\le \exp\{10^{188}(\log(8d))^2(\log\log(8d))^2\}+16^{4\log(240\log(4H))}.\end{aligned}$$

Let k be a non-negative integer. Set $m_0=T^k$ and $m_1=T^{k+1}$. Our choice for T contradicts (3.3). This shows that the assumptions of Lemma 1 cannot be satisfied. Consequently, there exists an integer n_k with $T^k\le n_k<T^{k+1}$ such that no block of n_k digits occurs twice in the prefix of length $3n_k$ of the b-ary expansion of ξ. This implies that

$$p(n_k,\xi,b)\ge 2n_k.$$

Let $n > n_0$ be a positive integer that does not belong to the sequence $(n_k)_{k\geq 0}$. Let k be the integer determined by the inequalities

$$T^k \leq n_k < n < n_{k+1} < T^{k+2}.$$

As proved in [10], since ξ is irrational, the function $m \mapsto p(m, \xi, b)$ is strictly increasing and

$$p(m, \xi, b) \geq m + 1, \quad \text{for } m \geq 1. \tag{3.4}$$

Consequently, the function $g : m \mapsto p(m, \xi, b) - m$ is non-decreasing (to see this, just compute $g(m + 1) - g(m)$), and we get

$$p(n, \xi, b) - n \geq p(n_k, \xi, b) - n_k \geq n_k \geq T^k \geq \frac{n}{T^2},$$

hence,

$$p(n, \xi, b) \geq \left(1 + \frac{1}{T^2}\right) n. \tag{3.5}$$

We infer from (3.4) that inequality (3.5) remains true for every n less than or equal to n_0. This concludes the proof of Theorem 2. □

Acknowledgements. I am pleased to thank Sergei Konyagin, whose question is at the origin of the present note, and Julien Cassaigne for stimulating discussion. Thanks are also due to the referee for his very careful reading.

References

[1] A. Aberkane, *Suites de complexité inférieure à* $2n$, Bull. Belg. Math. Soc. 8 (2001), 161–180.

[2] A. Aberkane, *Words whose complexity satisfies* $\lim p(n)/n = 1$, Theoret. Comput. Sci. 307 (2003), 31–46.

[3] B. Adamczewski and Y. Bugeaud, *On the complexity of algebraic numbers I. Expansions in integer bases*, Ann. of Math. 165 (2007), 547–565.

[4] B. Adamczewski and Y. Bugeaud, *Mesures de transcendance et aspects quantitatifs de la méthode de Thue–Siegel–Roth–Schmidt.* Preprint. `http://www-irma.u-strasbg.fr/~bugeaud/travaux/MesTransRevu.pdf`

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

[6] Y. Bugeaud and J.-H. Evertse, *On two notions of complexity of algebraic numbers.* Preprint. `http://front.math.ucdavis.edu/0709.1560`

[7] J.-H. Evertse and H.P. Schlickewei, *A quantitative version of the Absolute Subspace Theorem*, J. reine angew. Math. 548 (2002), 21–127.

[8] S. Ferenczi, *Rank and symbolic complexity*, Ergodic Th. Dyn. Systems 16 (1996), 663–682.

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

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

[11] W. M. Schmidt, Diophantine Approximation. Lecture Notes in Mathematics 785, Springer, 1980.

[12] W. M. Schmidt, *The subspace theorem in Diophantine approximation*, Compositio Math. 69 (1989), 121–173.

Yann Bugeaud, Université Louis Pasteur, U. F. R. de mathématiques
7, rue René Descartes, 67084 STRASBOURG Cedex (FRANCE)

bugeaud@math.u-strasbg.fr