Diophantine Analysis and Related Fields 2006 in honor of Prof. Iekata SHIOKAWA
7–10 March, 2006, Keio University, Yokohama, JAPAN
Conference Proceedings, Seminar on Mathematical Sciences, Keio University

Dedicated to Professor Iekata Shiokawa on his 65th birthday

DIOPHANTINE ANALYSIS AND WORDS

MICHEL WALDSCHMIDT

ABSTRACT. There is no explicitly known example of a triple (g, a, x), where $g \geq 3$ is an integer, a a digit in $\{0, \ldots, g-1\}$ and x a real algebraic irrational number, for which one can claim that the digit a occurs infinitely often in the g–ary expansion of x.

In 1909 and later in 1950, É. Borel considered such questions and suggested that the g–ary expansion of any algebraic irrational number in any base $g \geq 2$ satisfies some of the laws that are satisfied by almost all numbers. For instance, the frequency where a given finite sequence of digits occurs should depend only on the base and on the length of the sequence.

Hence there is a huge gap between the established theory and the expected state of the art. However, some progress have been made recently, mainly thanks to clever use of the Schmidt's subspace Theorem. We review some of these results.

CONTENTS

1. Introduction: three dimensional billiards 1
2. Complexity of words 3
2.1. Borel and normal numbers 3
2.2. Words 4
2.3. Finite automata and automatic sequences 4
2.4. Examples 5
2.5. BBP numbers 9
3. Words and transcendence 9
3.1. Number of 1's in the binary expansion of a real number 9
3.2. Complexity of the g-ary expansion of an algebraic number 10
3.3. Diophantine Approximation 11
3.4. Further transcendence results 12
4. Continued Fractions 13
4.1. Complexity of the continued fraction expansion of an algebraic number 13
4.2. The Fibonacci continued fraction 13
5. Open problems 14
References 14

1. INTRODUCTION: THREE DIMENSIONAL BILLIARDS

As a matter of introduction, we reproduce here three reviews in Mathematical Reviews from papers of which Iekata Shiokawa is a co-author: they deal with billiards and the complexity of words which we are interested in.

Date: September 4, 2006. Many thanks to Boris Adamczewski, Jean-Paul Allouche, Stéphane Fischler and Tanguy Rivoal for their useful comments on a preliminary version. I am also grateful to Takao Komatsu for his invitation to participate, for the organization of the conference as well as for editing the proceedings.

Review MR1279582 of [16]
Arnoux, Pierre(F-PARIS 7-M); Mauduit, Christian(F-LYON-MI); Shiokawa, Iekata(J-KEIOE); Tamura, Jun-ichi

Rauzy's conjecture on billiards in the cube.
Tokyo J. Math. **17** (1994), no. 1, 211–218.
58F03 (11B99)

The authors consider the game of billiards in the cube $[0,1]^3$. The two faces $x_i = 0$ and $x_i = 1$ are labelled i $(i = 1, 2, 3)$. Consider a particle which reflects on the six sides of the cube. Its orbit can be coded as an infinite sequence on the alphabet $\{1, 2, 3\}$ which records the consecutive sides that the particle hits. Suppose that the initial direction $(\alpha_1, \alpha_2, \alpha_3)$ is such that α_1, α_2, α_3 are $\mathbf{Q}$-linearly independent, and suppose furthermore that the forward orbit avoids the edges of the cube. The authors establish a conjecture of Rauzy according to which the number of distinct factors of length n in the coded orbit is $p(n) = n^2 + n + 1$. The authors prove this result in an elementary way with no reference to ergodic techniques.

Reviewed by M. Mendès France

Review MR1259106 of [15]
Arnoux, Pierre(F-PROVS-DM); Mauduit, Christian(F-LYON-LD); Shiokawa, Iekata(J-KEIOE); Tamura, Jun-ichi

Complexity of sequences defined by billiard in the cube.
Bull. Soc. Math. France **122** (1994), no. 1, 1–12.
11B85 (05B45 58F03)

The billiard problem in a square is well known. Consider the orbit of a billiard ball which one can code in the following way: each time the ball meets a "vertical" side [resp. "horizontal"] mark 1 [resp. 2]. An orbit is thus represented by an infinite sequence of 1's and 2's. Let $p(n)$ be the complexity of the sequence, i.e., the number of words of length n that occur in the sequence. A classical result states that either $p(n)$ is uniformly bounded if the initial direction (α, β) is rational (i.e., $\alpha/\beta \in \mathbf{Q} \cup \{\infty\}$) and the orbit is periodic, or $p(n) = n + 1$ if $\alpha/\beta \in \mathbf{R} \setminus \mathbf{Q}$ (the sequence is Sturmian).

The authors discuss the similar 3-dimensional problem, i.e. the billiard problem in the cube. An orbit is now coded by an infinite sequence on three symbols. Let $(\alpha_1, \alpha_2, \alpha_3)$ be the initial direction. Suppose $\alpha_1, \alpha_2, \alpha_3$ are $\mathbf{Q}$-independent. The authors prove the beautiful result $p(n) = n^2 + n + 1$. The proof is by no means easy: it combines skillful geometry on the torus with combinatorics.

The authors conclude the paper with questions pertaining to higher dimensions. Let $P(n, s)$ be the complexity of the orbit-sequence in the s-dimensional cube. It is not known whether $P(n, s)$ depends or not on the initial direction $(\alpha_1, \alpha_2, \dots, \alpha_s)$ even if these numbers are $\mathbf{Q}$-independent. The authors show that if $\min\{n, s\} \le 2$ then $P(n, s)$ exists and $P(n, s) = P(s, n)$. We are left with an intriguing problem: Do these results hold for all n, s?

Reviewed by M. Mendès France

Review MR1193183 of [38]
Shiokawa, Iekata(J-KEIO); Tamura, Jun-ichi

Description of sequences defined by billiards in the cube.
Proc. Japan Acad. Ser. A Math. Sci. **68** (1992), no. 7, 207–211.
11K55

In this paper "billiard" sequences in the unit cube $[0,1]^3$ are considered. Starting at a point P and moving with constant velocity $(1, \alpha, \beta)$, a particle is reflected at the faces of the cube. A sequence $w = (w_n)$ with entries x_i $(i = 1, 2, 3)$ is defined such that $w_n = x_i$ if the nth reflection is caused by a face which is orthogonal to the x_i-axis. The subword complexity of such sequences (and related ones) was extensively studied by Rauzy, Arnoux, Mauduit and others. The present authors present a precise description of such sequences in terms of the partial quotients of the simple continued fractions of $\alpha, \beta, \alpha/\beta$ and the digits appearing in some related expansions provided that $\alpha, \beta, \alpha/\beta$ are irrational $(1 > \alpha > \beta > 0)$ and the initial point P is "lattice-free" with respect to v (i.e., the path of the particle never touches the edges of the cube).

Reviewed by Robert F. Tichy

2. Complexity of words

2.1. **Borel and normal numbers.** In two papers, the first one published in 1909 [24] and the second one in 1950 [25], É. Borel considered the g–ary expansion of an algebraic irrational real number, where $g \geq 2$ is a positive integer. He suggested that this expansion should satisfy some of the laws shared by almost all numbers (for Lebesgue's measure).

Let $g \geq 2$ be an integer. Any real number x has a unique expansion

$$x = a_{-k}g^k + \cdots + a_{-1}g + a_0 + a_1g^{-1} + a_2g^{-2} + \cdots$$

where $k \geq 0$ is an integer and the a_i for $i \geq -k$, namely the digits of x in the expansion in base g of x, belong to the set $\{0, 1, \dots, g-1\}$. Unicity is subject to the condition that the sequence $(a_i)_{i\geq -k}$ is not ultimately constant equal to $g-1$. We write this expansion

$$x = a_{-k} \cdots a_{-1}a_0.a_1a_2 \cdots$$

Example. In base 10 (*decimal expansion*):

$$\sqrt{2} = 1.4142135623730950488016887242\,0 \dots$$

and in base 2 (*binary expansion*):

$$\sqrt{2} = 1.0110101000001001111001100110011111110011101111001100100100001000101 \dots$$

The first question in this direction is whether each digit always occurs at least once.

Conjecture 2.1. *Let x be an irrational algebraic real number, $g \geq 3$ a positive integer and a an integer in the range $0 \leq a \leq g-1$. Then the digit a occurs at least once in the g–ary expansion of x.*

There is no explicitly known example of a triple (g, a, x), where $g \geq 3$ is an integer, a a digit in $\{0, \dots, g-1\}$ and x a real algebraic irrational number, for which one can claim that the digit a occurs infinitely often in the g–ary expansion of x. Another open problem is to produce an explicit pair (x, g) where $g \geq 3$ is an integer and x a real algebraic irrational number, for which we can claim that the number of digits which occur infinitely many times in the g–ary expansion of x is at least 3. Even though few results are known, our knowledge is not completely vacuous, even if explicit examples are lacking: according to K. Mahler (see Theorem M in [4]), *for any $g \geq 2$ and any $n \geq 1$ there exist algebraic irrational numbers such that any block of n digits occurs infinitely often in the g–ary expansion of ξ.*

If a real number x satisfies Conjecture 2.1 for all g and a, then it follows that for any g each given sequence of digits occurs infinitely often in the g–ary expansion of x. This is easy to see by considering powers of g. For instance, Conjecture 2.1 with $g = 4$ implies that each of the four sequences $(0, 0)$, $(0, 1)$, $(1, 0)$, $(1, 1)$ should occur infinitely often in the binary expansion of any irrational algebraic real number x.

Borel asked more precise questions on the frequency of occurrences of sequences of binary digits of real irrational algebraic numbers. We need to introduce some definitions.

Firstly, a real number x is called *simply normal in base g* if each digit occurs with frequency $1/g$ in its g–ary expansion.

Secondly, a real number x is called *normal in base g* or *g–normal* if it is simply normal in base g^m for any $m \geq 1$. Hence a real number x is normal in base g if and only if, for any $m \geq 1$, each sequence of m digits occurs with frequency $1/g^m$ in its g–ary expansion.

Finally, a number is called *normal* if it is normal in any base $g \geq 2$.

Borel suggested that *each real irrational algebraic number should be normal.*

Conjecture 2.2 (É. Borel, 1950)**.** *Let x be an irrational algebraic real number and $g \geq 2$ a positive integer. Then x is normal in base g.*

Almost all numbers (for Lebesgue's measure) are normal, examples of computable normal numbers have been constructed (W. Sierpinski, H. Lebesgue, V. Becher and S. Figueira – see [22]), but the known algorithms to compute such examples are fairly complicated ("ridiculously exponential", according to [22]).

An example of a 2–normal number (Champernowne 1933, Bailey and Crandall 2001 [21]) is the *binary Champernowne number*, obtained by concatenation of the sequence of integers

$$0.\,1\,10\,11\,100\,101\,110\,111\,1000\,1001\,1010\,1011\,1100\,1101\,1110 \dots$$

A closed formula for this number is

$$\sum_{k\geq 1} k2^{-c_k} \quad \text{with} \quad c_k = k + \sum_{j=1}^{k} [\log_2 j].$$

Here is another example (Korobov, Stoneham ...– see [20]): *if a and g are coprime integers >1, then*

$$\sum_{n\geq 0} a^{-n} g^{-a^n}$$

is normal in base g.

A further example, due to A.H. Copeland and P. Erdős (1946), of a normal number in base 10 is

$$0.23571113171923\ \ldots$$

which is obtained by concatenation of the sequence of prime numbers.

2.2. **Words.** We recall some basic facts from language theory – see for instance [14, 34].

We consider an alphabet A with g letters. The free monoid A^* on A is the set of *finite words* $a_1 \ldots a_n$ where $n \geq 0$ and $a_i \in A$ for $1 \leq i \leq n$. The law on A^* is called *concatenation*.

The number of letters of a finite word is its *length*: the length of $a_1 \ldots a_n$ is n.

The number of words of length n is g^n for $n \geq 0$. The single word of length 0 is the empty word e with no letter. It is the neutral element for the concatenation.

We shall consider *infinite words* $w = a_1 \ldots a_n \ldots$ A *factor of length* m of such a w is a word of the form $a_k a_{k+1} \ldots a_{k+m-1}$ for some $k \geq 1$.

The *complexity* of an infinite word w is the function $p(m)$ which counts, for each $m \geq 1$, the number of distinct factors of w of length m. Hence for an alphabet A with g elements we have $1 \leq p(m) \leq g^m$ and the function $m \mapsto p(m)$ is non-decreasing. Conjecture 2.1 is equivalent to the assertion that the complexity of the sequence of digits in base g of an irrational algebraic number should be $p(m) = g^m$.

An infinite word is periodic if and only if its complexity is bounded. If the complexity $p(m)$ of a word satisfies $p(m+1) = p(m)$ for one value of m, then $p(m+k) = p(m)$ for any $k \geq 0$, hence the word is periodic. It follows that the complexity of a non-periodic word satisfies $p(m) \geq m+1$. Following Morse and Hedlund (1938), a word of minimal complexity $p(m) = m+1$ is called a *Sturmian word*. Sturmian words are those which encode with two letters the orbits of square billiard starting with an irrational angle. It is easy to check that on the alphabet $\{a,\ b\}$, a Sturmian word w is characterized by the property that *for each $m \geq 1$, there is exactly one factor v of w of length m such that both va and vb are factors of w of length $m+1$.*

Let A and B be two finite sets. A map from A to B^* can be uniquely extended to a homomorphism between the free monoids A^* and B^*. We call *morphism from A to B* such a homomorphism. The morphism is *uniform* if all words in the image of A have the same length.

Let ϕ be a morphism from A into itself. Assume that there exists a letter a such that $\phi(a) = au$, where u is a non-empty word satisfying $\phi^k(u) \neq e$ for every $k \geq 0$. Then the sequence of finite words $(\phi^k(a))_{k\geq 1}$ converges in $A^{\mathbf{N}}$ (endowed with the product topology of the discrete topology on each copy of A) to an infinite word $w = au\phi(u)\phi^2(u)\phi^3(u)\ldots$. This infinite word is clearly a fixed point for ϕ and we say that w *is generated by the morphism ϕ.*

If, moreover, every letter occurring in w occurs at least twice, then we say that w is generated by a *recurrent morphism.*

If the alphabet A has two letters, then we say that w is generated by a *binary morphism.*

More generally, an infinite sequence w in $A^{\mathbf{N}}$ is said to be *morphic* if there exist a sequence u generated by a morphism defined over an alphabet B and a morphism ϕ from B to A such that $w = \phi(u)$.

2.3. **Finite automata and automatic sequences.** A formal definition of a finite automaton is given, for instance, in [14] § 4.1, [34] § 1.3.2, [10] § 3.3 and [2] § 3. It involves finite sets, namely the set of states i, a, $b\ldots$, the set of transitions: here we shall use only 0 or 1, and two sets of initial and terminal states.

We do not give the exact definition but we propose a number of examples below (§2.4).

Let $g \geq 2$ be an integer. An infinite sequence $(a_n)_{n\geq 0}$ is said to be *g–automatic* if a_n is a finite-state function of the representation of n in base g: this means that there exists a finite automaton starting with the g–ary expansion of n as input and producing the term a_n as output.

According to a theorem due to A. Cobham [31], automatic sequences have a complexity $p(m) = O(m)$ (cf. § 10.3 in [14]). Also he proved in [31] that automatic sequences are the same as uniform morphic sequences. See for instance § 6.3 in [14] and Theorem 4.1 in [8].

Automatic sequences are between periodicity and chaos. They occur in connection with harmonic analysis, ergodic theory, fractals, Feigenbaum tree, quasi-cristals, transition phases in statistical mechanics,... see [13, 28, 10] and Chap. 17 in [14].

2.4. **Examples.**

2.4.1. *The Fibonacci word.* Consider the alphabet $A = \{a, b\}$. Start with $f_1 = b$, $f_2 = a$ and define (concatenation): $f_n = f_{n-1}f_{n-2}$. Hence

$$f_3 = ab, \quad f_4 = aba, \quad f_5 = abaab, \quad f_6 = abaababa \quad f_7 = abaababaabaab, \dots$$

There is a unique word

$$w = abaababaabaababaababaabaababaabaab\dots$$

of which f_n is the prefix of length F_n (the Fibonacci number of index n) for any $n \geq 2$. This is the *Fibonacci word*; it is generated by a binary recurrent morphism [14] § 7.1: namely, it is the fixed point of the morphism $a \mapsto ab$, $b \mapsto a$: under this morphism, the image of f_n is f_{n+1}.

Proposition 2.3. *The Fibonacci word is Sturmian.*

The factors of w are given as follows:

$$\begin{array}{cccccccccc}
 & & & & & & & & aabaa & \dots \\
 & & & & & & & \nearrow & & \\
a & \to & aa & \to & aab & \to & aaba & \to & aabab & \dots \\
 & \searrow & & & & & & & & \\
 & & ab & \to & aba & \to & abaa & \to & abaab & \dots \\
 & & & & & \searrow & & & & \\
 & & & & & & abab & \to & ababa & \dots \\
b & \to & ba & \to & baa & \to & baab & \to & baaba & \dots \\
 & & & \searrow & & & & & & \\
 & & & & bab & \to & baba & \to & babaa & \dots
\end{array}$$

Sketch of the proof of Proposition 2.3.

We outline a proof of the fact that *for each $m \geq 1$, the number $p(m)$ of factors of w of length m in w is $m + 1$.*

The *first step* is to check that the Fibonacci word is not periodic. Indeed, the word f_n has length F_n, it consists of F_{n-1} letters a and F_{n-2} letters b. Hence the proportion of a in the Fibonacci word w is $1/\Phi$, where Φ is the Golden number

$$\Phi = \frac{1+\sqrt{5}}{2}$$

which is an irrational number.

Remark. The proportion of b in w is $1/\Phi^2$ with $(1/\Phi) + (1/\Phi^2) = 1$, as expected!

Here is the *second step* of the proof of Proposition 2.3: for $n \geq 3$ the word f_n can be written

$$f_n = \begin{cases} u_n ba & \text{for even } n, \\ u_n ab & \text{for odd } n, \end{cases}$$

with u_n of length $F_n - 2$. Let us check by induction that *the word u_n is palindromic.*

Indeed, for even n we have

$$u_{n+1} = f_n u_{n-1} = u_{n-1}abu_{n-2}bau_{n-1}$$

while for odd n the formula is

$$u_{n+1} = f_n u_{n-1} = u_{n-1}bau_{n-2}abu_{n-1}.$$

Now the *third step*: we claim that

$$f_{n-1}f_n = \begin{cases} u_{n+1}ba & \text{for even } n, \\ u_{n+1}ab & \text{for odd } n, \end{cases}$$

which means that *in the two words $f_{n-1}f_n$ and f_{n+1}, only the last two letters are not the same.*

We prove this claim as follows: write $f_{n-1}f_n$ as $f_{n-1}f_{n-1}f_{n-2}$; for even n we have $u_{n+1} = f_{n-1}u_{n-2}bau_{n-1}$ and

$$f_{n-1}f_n = f_{n-1}u_{n-1}abu_{n-2}ba = f_{n-1}u_{n-2}bau_{n-1}ba = u_{n+1}ba,$$

while for odd n we have $u_{n+1} = f_{n-1}u_{n-2}abu_{n-1}$ and

$$f_{n-1}f_n = f_{n-1}u_{n-1}bau_{n-2}ab = f_{n-1}u_{n-2}abu_{n-1}ab = u_{n+1}ab.$$

The *next and last step* in the proof of Proposition 2.3 is to check that the number of factors of length $F_n - 1$ in w is at most F_n. The proof involves the factorization of w on $\{f_{n-1}, f_{n-2}\}$. We leave the details to the reader.

It follows that $p(m) = m + 1$ for infinitely many m, hence for all $m \geq 1$. □

A result of L.V. Danilov (1972) concerning the sequence

$$(v_n)_{n\geq 0} = (0, 1, 0, 0, 1, 0, 1, 0, 0, \ldots)$$

derived from the Fibonacci word on the alphabet $\{0, 1\}$, is that, for any $g \geq 2$ the number

$$\sum_{n\geq 0} v_n g^{-n}$$

is transcendental.

Proposition 2.4. *The Fibonacci word is not automatic.*

Proposition 2.4 follows from a result of A. Cobham [31] according to which the frequency of a letter in an automatic word, if it exists, is a rational number.

The origin of the Fibonacci sequence is a model of growth of a population of rabbits. Denote a pair of young rabbits by 0 and a pair of adults by 1. From one year to the next one, the young pair becomes adult, which we write as $0 \to 1$, while the adult pair stays alive and produces a young pair: $1 \to 10$. This yields to the dynamical system

$$0 \to 1 \to 10 \to 101 \to 10110 \to 10110101 \to \ldots$$

and the sequence $(R_1, R_2, R_3, \ldots)$ of Fibonacci's rabbits

$$1,\ 0,\ 1,\ 1,\ 0,\ 1,\ 0, 1,\ 1,\ 0,\ 1,\ 1,\ 0,\ 1,\ \ldots$$

Any integer $n \geq 2$ has a unique representation as sums of two Fibonacci numbers F_m, $m \geq 2$, with the property that no Fibonacci numbers with consecutive indices occur in the sum. This representation yields the following algorithm to decide whether R_n is 1 or 0. If the smallest index in the decomposition of n is even, then $R_n = 1$, if it is odd, then $R_n = 0$. For instance $51 = F_9 + F_7 + F_4 + F_2$, the smallest index, namely 2, is even, hence $R_{51} = 1$.

Denote by $\Phi = (1 + \sqrt{5})/2$ the Golden Number. The sequence of indices n such that $R_n = 1$ is

$$[\Phi] = 1,\ [2\Phi] = 3,\ [3\Phi] = 4,\ [4\Phi] = 6,\ \ldots$$

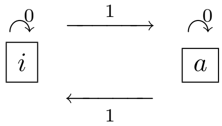
while the sequence of n for which $R_n = 0$ is

$$[\Phi^2] = 2,\ [2\Phi^2] = 5,\ [3\Phi^2] = 7,\ [4\Phi^2] = 8,\ \ldots$$

For instance $32\Phi = 51, 77\ \ldots$ hence $[32\Phi] = 51$ and $R_{51} = 1$.

This is an example of a *Beatty sequence*.

2.4.2. *The Prouhet–Thue–Morse word abbabaabbaababbab....* The finite automaton



with $f(i) = 0$, $f(a) = 1$

produces the sequence $a_0 a_1 a_2 \dots$ where, for instance, a_9 is $f(i) = 0$, since

$$1001[i] = 100[a] = 10[a] = 1[a] = i.$$

This is the *Prouhet–Thue–Morse sequence*

$$01101001100101101\dots,$$

where the $n+1$-th term a_n is 1 if the number of 1's (which is the same as the sum of the binary digits) in the binary expansion of n is odd, 0 if it is even (see [14], § 1.6).

If, in the Prouhet–Thue–Morse sequence $01101001100101101\dots$ we replace 0 by a and 1 by b, we obtain the *Prouhet–Thue–Morse word* on the alphabet $\{a, b\}$, which starts with

$$w = abbabaabbaababbab\dots$$

This word is generated by a binary recurrent morphism (see [14] § 6.2): it is the fixed point of the morphism $a \mapsto ab$, $b \mapsto ba$.

An interesting property of this sequence (due to A. Thue, 1906) is that, if w is a finite word and a a letter such that wwa is a factor of the Prouhet–Thue–Morse word, then a is not the first letter of w. Therefore, in the Prouhet–Thue–Morse word, no three consecutive identical blocks like 000 or 111 or 010101 or 101010 or 001001001 ... occurs.

The *Prouhet–Thue–Morse–Mahler number in base* $g \geq 2$ is the number

$$\xi_g = \sum_{n \geq 0} \frac{a_n}{g^n}$$

where $(a_n)_{n\geq 0}$ is the Prouhet–Thue–Morse sequence. The g–ary expansion of ξ_g starts with

$$0.1101001100101101\dots$$

These numbers were considered by K. Mahler, who proved in 1929 that they are transcendental [35].

The idea of proof is as follows (see [35] Example 1.3.1, where the complete proof is given). Consider the function

$$f(z) = \prod_{n\geq 0}(1 - z^{2^n}) \quad \text{which satisfies} \quad f(z) = \sum_{n \geq 0}(-1)^{a_n} z^n.$$

For $a \in \{0, 1\}$, we can write $(-1)^a = 1 - 2a$. Hence

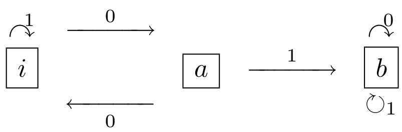
$$f(z) = \sum_{n\geq 0}(1 - 2a_n)z^n = \frac{1}{1-z} - 2\sum_{n\geq 0} a_n z^n.$$

Using the functional equation $f(z) = (1 - z)f(z^2)$, Mahler proves that $f(\alpha)$ is transcendental for any algebraic number α satisfying $0 < |\alpha| < 1$. □

2.4.3. *The Baum–Sweet sequence.* For $n \geq 0$ define $a_n = 1$ if the binary expansion of n contains no block of consecutive 0's of odd length, $a_n = 0$ otherwise: the *Baum–Sweet sequence* $(a_n)_{n\geq 0}$ starts with

$$1\,1\,0\,1\,1\,0\,0\,1\,0\,1\,0\,0\,1\,0\,0\,1\,1\,0\,0\,1\,0\dots$$

This sequence is automatic, associated with the automaton



with $f(i) = 1$, $f(a) = 0$, $f(b) = 0$. See [14], Example 5.1.7.

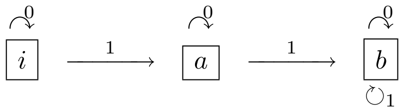
2.4.4. *Powers of* 2. The binary number

$$\xi := \sum_{n\geq 0} 2^{-2^n} = 0.1101000100000001000\cdots = 0.a_1a_2a_3\ldots$$

with

$$a_n = \begin{cases} 1 & \text{if } n \text{ is a power of } 2, \\ 0 & \text{otherwise} \end{cases}$$

is 2–automatic, given by the automaton



with $f(i) = 0$, $f(a) = 1$, $f(b) = 0$.

The associated infinite word

$$\mathbf{v} = v_1v_2\cdots v_n\cdots = bbabaaabaaaaaaabaaa\cdots,$$

where

$$v_n = \begin{cases} b & \text{if } n \text{ is a power of } 2, \\ a & \text{otherwise,} \end{cases}$$

has complexity $p(m)$ bounded by $2m$; the initial values are

$m =$	1	2	3	4	5	6	$\cdots$
$p(m) =$	2	4	6	7	9	11	$\cdots$

2.4.5. *The Rudin–Shapiro word.* For $n \geq 0$, define $r_n \in \{a,\, b\}$ as being equal to a (respectively b) if the number of occurrences of the pattern 11 in the binary representation of n is even (respectively odd). This produces the Rudin–Shapiro word $aaabaabaaaabbbab\ldots$.

Let σ be the morphism defined from the monoid B^* on the alphabet $B = \{1,2,3,4\}$ into B^* by: $\sigma(1) = 12$, $\sigma(2) = 13$, $\sigma(3) = 42$ and $\sigma(4) = 43$. Let

$$\mathbf{u} = 121312421213\ldots$$

be the fixed point of σ begining with 1 and let φ be the morphism defined from B^* to $\{a,b\}^*$ by: $\varphi(1) = aa$, $\varphi(2) = ab$ and $\varphi(3) = ba$, $\varphi(4) = bb$. Then the Rudin–Shapiro word is $\varphi(\mathbf{u})$, hence it is morphic.

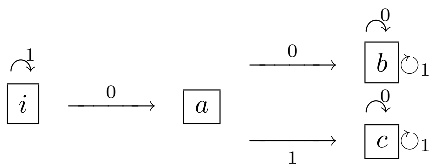
2.4.6. *Paper folding.* Folding a strip of paper always in the same direction, and then opening it up, yields a sequence of folds which can be encoded with 0 and 1. The resulting sequence $(u_n)_{n\geq 0}$:

$$1101100111001001\ \ldots$$

satisfies

$$u_{4n} = 1, \quad u_{4n+2} = 0, \quad u_{2n+1} = u_n$$

and is produced by the automaton



with $f(i) = f(a) = f(b) = 1$, $f(c) = 0$.

An equivalent definition for this sequence is given as follows (see [14] Example 5.1.6): the sequence $a_n = u_{n+1}$, $n \geq 1$, is defined recursively by $a_n = 1$ if n is a power of 2, say $n = 2^k$, and

$$a_{2^k+a} = 1 - a_{2^k-a} \quad \text{for} \quad 1 \leq a < 2^k.$$

For a connection between the paper folding sequence and the Prouhet–Thue–Morse sequence, see [17].

2.5. **BBP numbers.** An interesting approach towards Conjecture 2.2 is provided by *Hypothesis A* of Bailey and Crandall [21], who relate the question whether numbers like π, $\log 2$ and other constants are normal, to the following hypothesis involving the behaviour of the orbits of a discrete dynamical system.

Hypothesis A. *Let*

$$\theta := \sum_{n\geq 1} \frac{p(n)}{q(n)} \cdot g^{-n},$$

where $g \geq 2$ is a positive integer, $R = p/q \in \mathbf{Q}(X)$ a rational function with $q(n) \neq 0$ for $n \geq 1$ and $\deg p < \deg q$. Set $y_0 = 0$ and

$$y_{n+1} = gy_n + \frac{p(n)}{q(n)} \pmod 1$$

Then the sequence $(y_n)_{n\geq 1}$ either has finitely many limit points or is uniformly distributed modulo 1.

A connection with special values of G functions has been pointed out by J.C. Lagarias [32]. In this paper, Lagarias defines *BBP numbers*, referring to the paper [18] by D. Bailey, Jon Borwein and S. Plouffe, as numbers of the form

$$\sum_{n\geq 1} \frac{p(n)}{q(n)} \cdot g^{-n}$$

where $g \geq 2$ is an integer, p and q relatively prime polynomials in $\mathbf{Z}[X]$ with $q(n) \neq 0$ for $n \geq 1$.

For instance $\log 2$ is a BBP number in base 2 since

$$\sum_{n\geq 1} \frac{1}{n} \cdot x^n = -\log(1-x) \quad \text{and} \quad \sum_{n\geq 1} \frac{1}{n} \cdot 2^{-n} = \log 2.$$

Further, $\log 2$ is a BBP number in base $3^2 = 9$ since

$$\sum_{n\geq 1} \frac{1}{2n-1} \cdot x^{2n-1} = \frac{1}{2} \log \frac{1+x}{1-x}, \quad \text{hence} \quad \sum_{n\geq 1} \frac{6}{2n-1} \cdot 3^{-2n} = \log 2.$$

Furthermore, π^2 is a BBP number in base 2 and in base $3^4 = 81$ (D.J. Broadhurst 1999; see [19]).

3. Words and transcendence

3.1. **Number of 1's in the binary expansion of a real number.** Denote by $B(x, n)$ the number of 1's among the first n binary digits of an irrational real number x.

If x, y and $x + y$ are irrational, then

$$B(x + y, n) \leq B(x, n) + B(y, n) + 1.$$

If x, y and xy are irrational, then

$$B(xy, n) \leq B(x, n)B(y, n) + \log_2[x + y + 1].$$

If x is irrational, then for each integer $A > 0$ the bound

$$B(x, n)B(A/x, n) \geq n - 1 - \log_2[x + A/x + 1]$$

holds (see [20, 36]).

A consequence is that if a and b are two integers, both ≥ 2, then none of the powers of the transcendental number

$$\xi = \sum_{n\geq 1} a^{-b^n}$$

is simply normal in base 2. Also the lower bound

$$B(\sqrt{2}, n) \geq n^{1/2} + O(1)$$

can be deduced [36]. In [20], Theorem 7.1, D. Bailey, J. Borwein, R. Crandall and C. Pomerance obtain a similar lower bound valid for all algebraic irrational real numbers:

Theorem 3.1 (D. Bailey, J. Borwein, R. Crandall, C. Pomerance, 2004)**.** *Let x be a real algebraic number of degree $d \geq 2$. Then there is positive number C which depends only on x such that the number of* 1*'s among the first N digits in the binary expansion of x is at least $CN^{1/d}$.*

Further results related to Theorem 3.1 are given by T. Rivoal in [36] and by Y. Bugeaud in [27].

As pointed out by D. Bailey, J. Borwein, R. Crandall and C. Pomerance, it follows from Theorem 3.1 that for each $d \geq 2$, the number

$$\sum_{n \geq 0} 2^{-d^n}$$

is transcendental. The transcendence of the number

$$\sum_{n \geq 0} 2^{-2^n} \tag{1}$$

goes back to A. J. Kempner in 1916 ([14], § 13.10). A more general result, due to Mahler (1930, 1969; see [35], Theorem 1.1.2), is the transcendence of the values at algebraic points of the function $f(z) = \sum_{n \geq 0} z^{-d^n}$, for $d \geq 2$, which satisfies the functional equation $f(z^d) + z = f(z)$ for $|z| < 1$. As we shall see (§ 3) another proof rests on the approximation Theorem of Thue–Siegel–Roth–Ridout [1].

Another consequence of Theorem 3.1 is the transcendence of the number

$$\sum_{n \geq 0} 2^{-F_n},$$

whose binary digits are 1 at the Fibonacci indices 1, 2, 3, 5, 8, ... The transcendence of this number also follows from Mahler's method [20] as well as from the Theorem of Thue–Siegel–Roth–Ridout [1].

In 1968, A. Cobham conjectured that *automatic irrational numbers are transcendental.* J.H. Loxton and A.J. van der Poorten (1982, 1988 – see [14], § 13.10) tried to prove it, using Mahler's method. P.G. Becker in 1994 [23] (see also [2]) pointed out that Mahler's method yields only a weaker result so far: *for any given non-eventually periodic automatic sequence* $\mathbf{u} = (u_1, u_2, \dots)$*, the real number*

$$\sum_{k \geq 1} u_k g^{-k}$$

is transcendental, provided that the integer g is sufficiently large (in terms of $\mathbf{u}$*).*

It is a challenge to extend Mahler's method in order to prove Cobham's conjecture.

3.2. **Complexity of the g-ary expansion of an algebraic number.** The transcendence of a number whose sequence of digits is Sturmian has been proved by S. Ferenczi, C. Mauduit in 1997. It follows from their work (see [9]) that the complexity of the g–ary expansion of every irrational algebraic number satisfies

$$\liminf_{m \to \infty} (p(m) - m) = +\infty.$$

The main tool for the proof is a p–adic version of the Thue–Siegel–Roth Theorem due to Ridout (1957) – Theorem 3.7 below (see [4]).

Several papers have been devoted to the study of the complexity of the g-ary expansions of real algebraic numbers, in particular by J.-P. Allouche and L.Q. Zamboni (1998), R.N. Risley and L.Q. Zamboni (2000), B. Adamczewski and J. Cassaigne (2003). For a survey, see [4]. The main recent result is the following [2]:

Theorem 3.2 (B. Adamczewski, Y. Bugeaud, 2006)**.** *The complexity p of a real irrational algebraic number satisfies*

$$\liminf_{m \to \infty} \frac{p(m)}{m} = +\infty.$$

Corollary 3.3 (A. Cobham, 1968)**.** *If the sequence of digits of an irrational real number x is automatic, then x is transcendental.*

The main tool for the proof of Theorem 3.2 is a new, combinatorial transcendence criterion [6] obtained by B. Adamczewski, Y. Bugeaud and F. Luca as an application of Schmidt's subspace Theorem 3.8.

Theorem 3.2 implies the following statement related to the work of G. Christol (1979) [29], G. Christol, T. Kamae, M. Mendès France and G. Rauzy (1980) [30]:

Corollary 3.4. *Let p be a prime number, $g \geq p$ an integer and $(u_k)_{k\geq 1}$ a sequence of integers in the range $\{0, \dots, p-1\}$. The formal power series*

$$\sum_{k\geq 1} u_k X^k$$

and the real number

$$\sum_{k\geq 1} u_k g^{-k}$$

are both algebraic (over $\mathbf{F}_p(X)$ and over $\mathbf{Q}$, respectively) if and only if they are rational.

As an example (taken from § 6 in [11] and § 2.4 in [10]), consider the Prouhet–Thue–Morse sequence $(a_n)_{n\geq 0}$. The series

$$F(X) = \sum_{n\geq 0} a_n X^n$$

is algebraic over $\mathbf{F}_2(X)$, as it is a root of

$$(1+X)^3F^2 + (1+X)^2F + X = 0.$$

Hence Corollary 3.4 gives another proof of Mahler's transcendence result on the number

$$\sum_{n\geq 0} a_n g^{-n}.$$

3.3. **Diophantine Approximation.** Diophantine approximation theory enables one to prove that a number of the form

$$\sum_{n\geq 0} 2^{-u_n} \tag{2}$$

is transcendental, provided that the sequence $(u_n)_{n\geq 0}$ is increasing and grows sufficiently fast. The first statement in this direction goes back to J. Liouville in 1844.

Theorem 3.5 (J. Liouville, 1844)**.** *For any real algebraic number α there exists a constant $c > 0$ such that the set of $p/q \in \mathbf{Q}$ with $|\alpha - p/q| < q^{-c}$ is finite.*

Liouville's Theorem yields the transcendence of the value of a series like (2), provided that the increasing sequence $(u_n)_{n\geq 0}$ satisfies

$$\limsup_{n\to\infty} \frac{u_{n+1}}{u_n} = +\infty.$$

For instance $u_n = n!$ satisfies this condition: hence the number $\sum_{n\geq 0} 2^{-n!}$ is transcendental.

Theorem 3.6 (A. Thue, C.L. Siegel, K.F. Roth, 1950)**.** *For any real algebraic number α, for any $\epsilon > 0$, the set of $p/q \in \mathbf{Q}$ with $|\alpha - p/q| < q^{-2-\epsilon}$ is finite.*

Theorem 3.6 yields the transcendence of the series (2) under the weaker hypothesis

$$\limsup_{n\to\infty} \frac{u_{n+1}}{u_n} > 2.$$

The sequence $u_n = [2^{\theta n}]$ satisfies this condition as soon as $\theta > 1$. For example the transcendence of the number

$$\sum_{n\geq 0} 2^{-3^n}$$

follows from Theorem 3.6.

A stronger result follows from Ridout's Theorem 3.7 below, using the fact that the denominators 2^{u_n} are powers of 2: the condition

$$\limsup_{n\to\infty} \frac{u_{n+1}}{u_n} > 1$$

suffices to imply the transcendence of the sum of the series (2) – see [1].

An example is the transcendence of $\sum_{n\geq 0} 2^{-2^n}$ (see (1) above).

Theorem 3.7 (D. Ridout, 1957)**.** *For any real algebraic number α, for any $\epsilon > 0$, the set of $p/q \in \mathbf{Q}$ with $q = 2^k$ and $|\alpha - p/q| < q^{-1-\epsilon}$ is finite.*

The theorems of Thue–Siegel–Roth and Ridout are very special cases of Schmidt's subspace Theorem (1972) together with its p-adic extension by H.P. Schlickewei (1976). We state only a simplified version.

For $\mathbf{x} = (x_0, \dots, x_{m-1}) \in \mathbf{Z}^m$, define $|\mathbf{x}| = \max\{|x_0|, \dots, |x_{m-1}|\}$.

Theorem 3.8 (Schmidt's Subspace Theorem)**.** *Let $m \geq 2$ be a positive integer, S a finite set of places of $\mathbf{Q}$ containing the infinite place. For each $v \in S$ let $L_{0,v}, \dots, L_{m-1,v}$ be m independent linear forms in m variables with algebraic coefficients in the completion of $\mathbf{Q}$ at v. Let $\epsilon > 0$. Then the set of $\mathbf{x} = (x_0, \dots, x_{m-1}) \in \mathbf{Z}^m$ such that*

$$\prod_{v \in S} |L_{0,v}(\mathbf{x}) \cdots L_{m-1,v}(\mathbf{x})|_v \leq |\mathbf{x}|^{-\epsilon}$$

is contained in the union of finitely many proper subspaces of $\mathbf{Q}^m$.

Thue–Siegel–Roth's Theorem 3.6 follows from Theorem 3.8 by taking

$$S = \{\infty\}, \quad m = 2, \quad L_0(x_0, x_1) = x_0, \quad L_1(x_0, x_1) = \alpha x_0 - x_1.$$

Also Ridout's Theorem 3.7 is a consequence of Schmidt's subspace Theorem: in Theorem 3.8 take

$$m = 2, \quad S = \{\infty,\ 2\}, \quad L_{0,\infty}(x_0, x_1) = L_{0,2}(x_0, x_1) = x_0, \quad L_{1,\infty}(x_0, x_1) = \alpha x_0 - x_1, \quad L_{1,2}(x_0, x_1) = x_1.$$

For $(x_0, x_1) = (q, p)$ with $q = 2^k$, we have

$$|L_{0,\infty}(x_0, x_1)|_\infty = q, \quad |L_{1,\infty}(x_0, x_1)|_\infty = |q\alpha - p|, \quad |L_{0,2}(x_0, x_1)|_2 = q^{-1}, \quad |L_{1,2}(x_0, x_1)|_2 = |p|_2 \leq 1.$$

□

3.4. **Further transcendence results.** The previous results can be made effective in order to reach irrationality measures or transcendence measures for automatic numbers.

In 2006, B. Adamczewski and J. Cassaigne [8] solved a Conjecture of J. Shallit (1999) by proving that *a Liouville number cannot be generated by a finite automaton.*

They obtained irrationality measures for automatic numbers. Recall that the *irrationality exponent* of an irrational real number $x \in \mathbf{R} \setminus \mathbf{Q}$ is the least upper bound of the set of numbers κ for which the inequality

$$\left| x - \frac{p}{q} \right| < \frac{1}{q^\kappa}$$

has infinitely many solutions p/q.

For instance a Liouville number is a number whose irrationality exponent is infinite, while an irrational algebraic real number has irrationality exponent 2 (by Theorem 3.6), as do almost all real numbers.

Theorem 2.2 of [8] gives an explicit upper bound for the irrationality exponent for automatic irrational numbers. For the Prouhet–Thue–Morse–Mahler numbers for instance, the exponent of irrationality is ≤ 5. However there is no uniform upper bound for such exponents, as was pointed out to me by B. Adamczewski: the irrationality exponent for the automatic number associated with $\sigma(0) = 0^n 1$ and $\sigma(1) = 1^n 0$ is at least n.

Recently, B. Adamczewski and Y. Bugeaud obtained transcendence measures for automatic numbers: they proved that *automatic irrational numbers are either S or T–numbers* in Mahler's classification of transcendental numbers – cf [26]. This is a partial answer to a conjecture of P.G. Becker [8], according to which all automatic irrational numbers are S–numbers.

In another direction (see [2]), one deduces further transcendence results from Nesterenko 1996 work on the transcendence of values of theta series at rational points involving modular functions. For instance the transcendence of the number $\displaystyle\sum_{n \geq 0} 2^{-n^2}$ has been proved by D. Bertrand (1997), D. Duverney, K. Nishioka, K. Nishioka and I. Shiokawa (1998).

Another example from [2] is related to the word

$$\mathbf{u} = 012122122212222122222122222212222\dots$$

generated by the non-recurrent morphism $0 \mapsto 012$, $1 \mapsto 12$, $2 \mapsto 2$: *the number* $\eta = \sum_{k\geq 1} u_k 3^{-k}$ *is transcendental.*

4. Continued Fractions

We discussed above some Diophantine problems which are related with the g–ary expansion of a real number. Similar questions arrise with the continued fraction expansion of real numbers.

4.1. Complexity of the continued fraction expansion of an algebraic number. In 1949, A.Ya. Khintchine asked the following question: *are the partial quotients of the continued fraction expansion of a non-quadratic irrational algebraic real number bounded?* So far no example is known: it is not yet ruled out neither that all these partial quotients would be bounded for any such x, nor that they are unbounded for any such x. The common expectation seems to be that they are never bounded. The situation in finite characteristic is quite different: in 1976, L.E. Baum and M.M. Sweet constructed a formal series which is cubic over the field $\mathbf{F}_2(X)$, the continued fraction of which has partial quotients of bounded degree.

The very first transcendence results for numbers given by their continued fraction expansions are due to J. Liouville in 1844. This topic has been extensively developed by E. Maillet (1906) and later O. Perron (1929). In 1955, the Thue–Siegel–Roth Theorem enabled H. Davenport and K.F. Roth to prove deeper results. Further investigations are due to A. Baker (1962, 1964). The approximation results on real numbers by quadratic numbers due to W.M. Schmidt (1967) were a main tool for the next steps by J.L. Davison (1989) and by M. Queffélec in 1998 who established the transcendence of the Prouhet–Thue–Morse continued fraction. We mention also some work by P. Liardet and P. Stambul in 2000 and by J.-P. Allouche, J.L. Davison, M Queffélec and L.Q. Zamboni in 2001 where they prove the transcendence of Sturmian or morphic continued fractions. There are further papers by C. Baxa in 2004 and by B. Adamczewski, Y. Bugeaud and J.L. Davison (2005) who proved the transcendence of the Rudin–Shapiro and of the Baum–Sweet continued fractions.

Finally, in 2005 [3], B. Adamczewski and Y. Bugeaud proved that *the continued fraction expansion of an algebraic number of degree at least three cannot be generated by a binary morphism.*

4.2. The Fibonacci continued fraction. The Fibonacci word (recall §2.4.1) enabled D. Roy in [37] to construct transcendental real numbers ξ such that ξ and ξ^2 are surprisingly well simultaneously and uniformly approximated by rational numbers.

Recall that Φ denotes the Golden number, so that $\Phi^{-1} = \Phi - 1 = (-1+\sqrt{5})/2 = 0.618\dots$

Theorem 4.1 (D. Roy, 2003)**.** *Let A and B be two distinct positive integers. Let $\xi \in (0,1)$ be the real number whose continued fraction expansion is obtained from the Fibonacci word w by replacing the letters a and b by A and B:*

$$[0; A, B, A, A, B, A, B, A, A, B, A, A, B, A, B, A, A, B, A, B, A, A, B, A, A, B, A, B, A, A, B, A, A, B\dots]$$

Then there exists $c > 0$ such that the inequalities

$$0 < x_0 \leq X, \quad |x_0\xi - x_1| \leq cX^{-\Phi^{-1}}, \quad |x_0\xi^2 - x_2| \leq cX^{-\Phi^{-1}},$$

have a solution in $\mathbf{Z}^3$ for any sufficiently large value of X.

The arguments of D. Roy provides simplified proofs of Queffélec's results on the transcendence of the Thue–Morse and Fibonacci numbers.

Using Theorem 4.1 together with ideas of Davenport and Schmidt involving a transference theorem (Mahler's convex bodies), D. Roy also produced transcendental numbers which are surprisingly badly approximated by cubic algebraic integers.

Further results on Diophantine approximation of Sturmian continued fractions and on the simultaneous approximation of a number and its square have been worked out by Y. Bugeaud and M. Laurent, S. Fischler and more recently by D. Roy again.

For further references on Diophantine approximation, we refer to Bugeaud's recent book [26].

5. Open problems

Many problems related to the present topic are open, to trace their source would require more thorough bibliographical investigation. We give references to recent papers where they are quoted, the origin of these questions is most often much older.

Among the questions raised in [3] is the following one:

Does there exist an algebraic number of degree at least three whose continued fraction expansion is generated by a morphism ?

In the same vein the next question is open: *Do there exist an integer $g \geq 3$ and an algebraic number of degree at least three whose expansion in base g is generated by a morphism ?*

Other problems suggested by T. Rivoal [36] are:

- Let $g \geq 2$ be an integer. Give an explicit example of a real number $x > 0$ which is simply normal in base g and such that $1/x$ is not simply normal in base g.
- Same question with *normal in base g* in place of *simply normal in base g*.
- Same question with *normal* in place of *simply normal in base g*.

Remark. In [33], there is a construction of an automatic number, the inverse of which is not automatic. This answers by anticipation Problem 2 of § 13.9 in [14].

From the open problems in § 13.9 quoted in [14] we select the two following ones.

- Show that the number
$$\log 2 = \sum_{n \geq 1} \frac{1}{n} 2^{-n}$$
is not 2-automatic.
- Show that the number
$$\pi = \sum_{n \geq 0} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right) 2^{-4n}$$
is not 2-automatic.

One last open problem attributed to Mahler (see for instance [5]). Let $(e_n)_{n \geq 1}$ be an infinite sequence over $\{0, 1\}$ that is not ultimately periodic. Prove or disprove: *at least one of the two numbers*
$$\sum_{n \geq 1} e_n 2^{-n}, \quad \sum_{n \geq 1} e_n 3^{-n}$$
is transcendental.

From Conjecture 2.1 with $g = 3$ and $a = 2$, it follows that the second number should be always transcendental.

References

[1] B. Adamczewski – "Transcendance "à la Liouville" de certains nombres réels", *C. R. Math. Acad. Sci. Paris* **338** (2004), no. 7, p. 511–514.

[2] B. Adamczewski & Y. Bugeaud – "On the complexity of algebraic numbers I. Expansion in integer bases", to appear in Annals of Math., 2006.

[3] ———, "On the complexity of algebraic numbers II. Continued fractions", *Acta Math.* **195** (2005), p. 1–20.

[4] ———, "On the decimal expansion of algebraic numbers", *Fiz. Mat. Fak. Moksl. Semin. Darb.* **8** (2005), p. 5–13.

[5] ———, "Real and p-adic expansions involving symmetric patterns", to appear in Int. Math. Res. Not., 2006.

[6] B. Adamczewski, Y. Bugeaud & F. Luca – "Sur la complexité des nombres algébriques", *C. R. Math. Acad. Sci. Paris* **339** (2004), no. 1, p. 11–14.

[7] B. Adamczewski & J. Cassaigne – "On the transcendence of real numbers with a regular expansion", *J. Number Theory* **103** (2003), no. 1, p. 27–37.

[8] ———, "Diophantine properties of real numbers generated by finite automata", to appear in Compositio Math., http://arXiv.org/abs/math/0601604, 2006.

[9] J.-P. Allouche – "Nouveaux résultats de transcendance de réels à développement non aléatoire", *Gaz. Math.* (2000), no. 84, p. 19–34.

[10] ———, "Suites automatiques et séries formelles algébriques", *Dossier AF175, Mathématiques pour l'ingénieur, Techniques de l'ingénieur,* Octobre 2005, p. AF 175-1–AF 175-8.

[11] ———, "Automates et algébricités", *J. Théor. Nombres Bordeaux* **17** (2005), no. 1, p. 1–11.
[12] J.-P. Allouche, J. L. Davison, M. Queffélec & L. Q. Zamboni – "Transcendence of Sturmian or morphic continued fractions", *J. Number Theory* **91** (2001), no. 1, p. 39–66.
[13] J.-P. Allouche & M. Mignotte – "Arithmétique et automates", *Images des Mathématiques* (1988), p. 5–9, Courrier du CNRS, supplément au numéro **69**.
[14] J.-P. Allouche & J. Shallit – *Automatic sequences*, Cambridge University Press, Cambridge, 2003, Theory, applications, generalizations.
[15] P. Arnoux, C. Mauduit, I. Shiokawa & J.-i. Tamura – "Complexity of sequences defined by billiard in the cube", *Bull. Soc. Math. France* **122** (1994), no. 1, p. 1–12.
[16] ———, "Rauzy's conjecture on billiards in the cube", *Tokyo J. Math.* **17** (1994), no. 1, p. 211–218.
[17] R. Bacher – "La suite de Thue-Morse et la catégorie Rec", *C. R. Math. Acad. Sci. Paris* **342** (2006), no. 3, p. 161–164.
[18] D. Bailey, P. Borwein & S. Plouffe – "On the rapid computation of various polylogarithmic constants", *Math. Comp.* **66** (1997), no. 218, p. 903–913.
[19] D. H. Bailey & J. M. Borwein – "Experimental mathematics: recent developments and future outlook", in *Mathematics unlimited—2001 and beyond*, Springer, Berlin, 2001, p. 51–66.
[20] D. H. Bailey, J. M. Borwein, R. E. Crandall & C. Pomerance – "On the binary expansions of algebraic numbers", *J. Théor. Nombres Bordeaux* **16** (2004), no. 3, p. 487–518.
[21] D. H. Bailey & R. E. Crandall – "On the random character of fundamental constant expansions", *Experiment. Math.* **10** (2001), no. 2, p. 175–190.
[22] V. Becher & S. Figueira – "An example of a computable absolutely normal number", *Theoret. Comput. Sci.* **270** (2002), p. 947–958.
[23] P.-G. Becker – "k-regular power series and Mahler-type functional equations", *J. Number Theory* **49** (1994), no. 3, p. 269–286.
[24] É. Borel – "Les probabilités dénombrables et leurs applications arithmétiques", *Palermo Rend.* **27** (1909), p. 247–271.
[25] ———, "Sur les chiffres décimaux de $\sqrt{2}$ et divers problèmes de probabilités en chaînes", *C. R. Acad. Sci., Paris* **230** (1950), p. 591–593.
[26] Y. Bugeaud – *Approximation by algebraic numbers*, Cambridge Tracts in Mathematics, vol. 160, Cambridge University Press, Cambridge, 2004.
[27] ———, "On the b-ary expansion of an algebraic number", manuscript, 2006.
[28] R. Cerf – "Le modèle d'Ising et la coexistence des phases", *Images des Mathématiques* (2004), p. 47–51, Courrier du CNRS.
[29] G. Christol – "Ensembles presque périodiques k-reconnaissables", *Theoret. Comput. Sci.* **9** (1979), no. 1, p. 141–145.
[30] G. Christol, T. Kamae, M. Mendès France & G. Rauzy – "Suites algébriques, automates et substitutions", *Bull. Soc. Math. France* **108** (1980), no. 4, p. 401–419.
[31] A. Cobham – "Uniform tag sequences", *Math. Systems Theory* **6** (1972), p. 164–192.
[32] J. C. Lagarias – "On the normality of arithmetical constants", *Experiment. Math.* **10** (2001), no. 3, p. 355–368.
[33] S. Lehr, J. Shallit & J. Tromp – "On the vector space of the automatic reals", *Theoret. Comput. Sci.* **163** (1996), no. 1-2, p. 193–210.
[34] M. Lothaire – *Algebraic combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 90, Cambridge University Press, Cambridge, 2002.
[35] K. Nishioka – *Mahler functions and transcendence*, Lecture Notes in Mathematics, vol. 1631, Springer-Verlag, Berlin, 1996.
[36] T. Rivoal – "On the bits counting function of real numbers", manuscript; http://www-fourier.ujf-grenoble.fr/∼rivoal/articles/binreal.pdf, 2006.
[37] D. Roy – "Approximation simultanée d'un nombre et de son carré", *C. R. Math. Acad. Sci. Paris* **336** (2003), no. 1, p. 1–6.
[38] I. Shiokawa & J.-i. Tamura – "Description of sequences defined by billiards in the cube", *Proc. Japan Acad. Ser. A Math. Sci.* **68** (1992), no. 7, p. 207–211.

Université Pierre et Marie Curie–Paris 6, UMR 7586 IMJ Institut de Mathématiques de Jussieu, 175 rue du Chevaleret, Paris, F–75013 France
E-mail address: `miw@math.jussieu.fr`