31

WORDS AND TRANSCENDENCE

MICHEL WALDSCHMIDT

Dedicated to the 80th birthday of Professor K.F. Roth

Abstract. Is it possible to distinguish algebraic from transcendental real numbers by considering the b-ary expansion in some base $b \geqslant 2$? In 1950, É. Borel suggested that the answer is no and that for any real irrational algebraic number x and for any base $g \geqslant 2$, the g-ary expansion of x should satisfy some of the laws that are shared 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.

We are very far from such a goal: there is no explicitly known example of a triple (g, a, x), where $g \geqslant 3$ is an integer, a a digit in $\{0, \ldots, g-1\}$ and x a real irrational algebraic number, for which one can claim that the digit a occurs infinitely often in the g-ary expansion of x.

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

1. Normal Numbers and Expansion of Fundamental Constants

1.1. **Borel and Normal Numbers.** In two papers, the first [28] published in 1909 and the second [29] in 1950, Borel studied the g-ary expansion of real numbers, where $g \geqslant 2$ is a positive integer. In his second paper, he suggested that this expansion for a real irrational algebraic number should satisfy some of the laws shared by almost all numbers, in the sense of Lebesgue measure.

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

$$x = a_{-k}g^k + \cdots + a_{-1}g + a_0 + a_1 g^{-1} + a_2 g^{-2} + \cdots,$$

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

$$x = a_{-k} \cdots a_{-1} a_0 . a_1 a_2 \cdots .$$

Example. We have

$$\sqrt{2} = 1.41421356237309504880168872420\ldots$$

in base 10 (*decimal expansion*), whereas

$$\sqrt{2} = 1.0110101000001001111001100110011111111001110111001 1\ldots$$

in base 2 (*binary expansion*).

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

Conjecture 1.1. *Let x be an real irrational algebraic number, $g \geqslant 3$ a positive integer and a an integer in the range $0 \leqslant a \leqslant g-1$. Then the digit a occurs at least once in the g-ary expansion of x.*

For $g = 2$ it is obvious that each of the two digits 0 and 1 occurs infinitely often in the binary expansion of an irrational algebraic number. The same is true for each of the sequences of two digits 01 and 10. Conjecture 1.1 implies that each of the sequences of digits 00 and 11 should also occur infinitely often in such an expansion; apply Conjecture 1.1 with $g = 4$. There is no explicitly known example of a triple (g, a, x), where $g \geqslant 3$ is an integer, a a digit in $\{0, \ldots, g-1\}$ and x a real irrational algebraic 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 $g \geqslant 3$ is an integer and x a real irrational algebraic 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 and explicit examples are lacking, something is known. For any $g \geqslant 2$ and any $k \geqslant 1$ there exist real algebraic numbers x such that any sequence of k digits occurs infinitely often in the g-ary expansion of x. However the connection with algebraicity is weak: indeed Mahler proved in 1973 more precisely that for any real irrational number α and any sequence of k digits in the set $\{0, \ldots, g-1\}$, there exists an integer m for which the g-ary expansion of $m\alpha$ contains infinitely many times the given sequence; see [3, Theorem M] and [50]. According to Mahler, the smallest such m is bounded by g^{2k+1}. This estimate has been improved by Berend and Boshernitzan [26] to $2g^{k+1}$ and one cannot get better than $g^k - 1$.

If a real number x satisfies Conjecture 1.1 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.

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.

First, a real number x is called *simply normal in base g* if each digit occurs with frequency $1/g$ in its g-ary expansion. A very simple example in base 10 is

$$x = 0.123456789012345678901234567890\ldots,$$

where the sequence 1234567890 is repeated periodically, but this number is rational. We have

$$x = \frac{1\,234\,567\,890}{9\,999\,999\,999} = \frac{137\,174\,210}{1\,111\,111\,111}.$$

Second, a real number x is called *normal in base g* or *g-normal* if it is simply normal in base g^m for all $m \geqslant 1$. Hence a real number x is normal in base g if and only if, for all $m \geqslant 1$, 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 all bases $g \geqslant 2$.

Borel suggested in 1950 that each real irrational algebraic number should be normal.

Conjecture 1.2 (Borel, 1950). *Let x be a real irrational algebraic number and $g \geqslant 2$ a positive integer. Then x is normal in base g.*

As shown by Borel [28], almost all numbers, in the sense of Lebesgue measure, are normal. Examples of computable normal numbers have been constructed by Sierpinski, Lebesgue, Becher and Figueira; see [24]. However, the known algorithms to compute such examples are fairly complicated; indeed, "ridiculously exponential", according to [24].

An example of a 2-normal number is the *binary Champernowne number*, obtained by concatenation of the sequence of integers

$$0.1\,10\,11\,100\,101\,110\,111\,1000\,1001\,1010\,1011\,1100\,1101\,1110\,\ldots;$$

see [35, 54, 22]. A closed formula for this number is

$$\sum_{k\geqslant 1} k2^{-c_k}, \quad c_k = k + \sum_{j=1}^{k} \lfloor \log_2 j \rfloor;$$

see [22, p. 183].

Another example is given by Korobov, Stoneham, and others: if a and g are coprime integers greater than 1, then

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

is normal in base g; see [20].

A further example, due to Copeland and Erdős in 1946, of a normal number in base 10 is

$$0.23571113171923\ldots,$$

obtained by concatenation of the sequence of prime numbers.

As pointed out to the author by Tanguy Rivoal, a definition of what it means for a number to have a *random sequence of digits* is given in [33], the first concrete example being *Chaitin's omega number* which is the *halting probability of a universal self-delimiting computer with null-free data*. This number, which is normal and transcendental, appears also to be a good candidate for being a non-period in the sense of Kontsevich and Zagier [43].

1.2. **BBP Numbers.** An interesting approach towards Conjecture 1.2 is provided by *Hypothesis A* of Bailey and Crandall [22], 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*

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

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

$$y_{n+1} = g y_n + \frac{p(n)}{q(n)} \mod 1.$$

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

A connection with special values of G-functions has been noted by Lagarias [44]. In his paper, Lagarias defines *BBP numbers*, with reference to the paper [21] by Bailey, Borwein and Plouffe, as numbers of the form

$$\sum_{n\geqslant 1} \frac{p(n)}{q(n)}\, g^{-n},$$

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

Here are a few examples from [19]. Since

$$\sum_{n\geqslant 1} \frac{1}{n}\, x^n = -\log(1-x) \quad \text{and} \quad \sum_{n\geqslant 1} \frac{1}{2n-1}\, x^{2n-1} = \frac{1}{2}\log\frac{1+x}{1-x},$$

it follows that $\log 2$ is a BBP number in base 2 as well as in base 9, since

$$\log 2 = \sum_{n\geqslant 1} \frac{1}{n}\, 2^{-n} = \sum_{n\geqslant 1} \frac{6}{2n-1}\, 3^{-2n}. \tag{1}$$

Also $\log 3$ is a BBP number in base 4, since

$$\log 3 = 2\log 2 + \log\frac{3}{4} = \sum_{n\geqslant 0} \frac{1}{2^{2n}}\, \frac{1}{2n+1};$$

and π is a BBP number in base 16, since

$$\pi = \sum_{n\geqslant 0} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6}\right) 2^{-4n}. \tag{2}$$

Further examples are π^2 in base 64 and in base 81, $(\log 2)^2$ in base 64, and $\zeta(3)$ in base $2^{12} = 4096$.

1.3. **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 positive and irrational, then for all sufficiently large n,

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

If x, y and xy are positive and irrational, then for all sufficiently large n,

$$B(xy,n) \leqslant B(x,n)B(y,n) + \log_2\lfloor x+y+1\rfloor.$$

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

$$B(x,n)B\left(\frac{A}{x},n\right) \geqslant n - 1 - \left\lfloor \log_2\left(x + \frac{A}{x} + 1\right)\right\rfloor$$

holds for all sufficiently large n; see [20, 56].

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

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

is simply normal in base 2. Also the lower bound $B(\sqrt{2},n) \geqslant n^{1/2} + O(1)$ can be deduced [56]. In [20, Theorem 7.1], Bailey, Borwein, Crandall and Pomerance have obtained a similar lower bound valid for all real irrational algebraic numbers.

Theorem 1.1 (Bailey, Borwein, Crandall, and Pomerance, 2004). *Let x be a real algebraic number of degree at least 2. Then there is a 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 1.1 are given by Rivoal [56], Bugeaud [31], and Bugeaud and Evertse [32].

As pointed out by Bailey, Borwein, Crandall and Pomerance, it follows from Theorem 1.1 that for each $d \geqslant 2$, the number

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

is transcendental. The transcendence of the number

$$\sum_{n\geqslant 0} 2^{-2^n} \tag{3}$$

goes back to Kempner in 1916; see [15, Section 13.10]. Other proofs are available, which rest either on Mahler's method, to be discussed in Section 3.1, or on the approximation theorem of Thue–Siegel–Roth–Ridout, to be discussed in Section 4; see also [61, Section 1.6], and [57, 58, 1].

Another consequence of Theorem 1.1 is the transcendence of the number

$$\sum_{n\geqslant 0} g^{-F_n}, \tag{4}$$

for any integer $g \geqslant 2$, where $(F_n)_{n\geqslant 0}$ is the Fibonacci sequence, with $F_0 = 0$, $F_1 = 1$ and $F_{n+1} = F_n + F_{n-1}$ for $n \geqslant 1$. Again, the transcendence of this number also follows from Mahler's method [20] as well as from the theorem of Thue–Siegel–Roth–Ridout [57, 58, 1].

2. Words and Automata

2.1. **Words.** We recall some basic facts from language theory; see for instance [15, 46].

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

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

The number of words of length n is g^n for $n \geqslant 0$. 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* $w = a_1a_2a_3 \dots$. A *factor of length* m of such a word w is a word of the form $a_ka_{k+1} \dots a_{k+m-1}$ for some $k \geqslant 1$.

The *complexity* of an infinite word w is the function $p(m)$ which counts, for each $m \geqslant 1$, the number of distinct factors of w of length m. Hence for an alphabet A with g elements we have $1 \leqslant p(m) \leqslant g^m$, and the function $m \mapsto p(m)$ is non-decreasing. Conjecture 1.1 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 all $k \geqslant 0$, hence the word is periodic. It follows that the complexity of a non-periodic word satisfies $p(m) \geqslant m + 1$. Following Morse and Hedlund, 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; see [16, 17, 62]. It is easy to check that on the alphabet $\{a, b\}$, a Sturmian word w is characterized by the property that for each $m \geqslant 1$, there is exactly one factor v of w of length m for which 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 such a homomorphism a *morphism from A to B*. 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 for which $\varphi(a) = au$, where u is a non-empty word satisfying $\varphi^k(u) \neq e$ for every $k \geqslant 0$. Then the sequence of finite words $(\varphi^k(a))_{k\geqslant 1}$ converges in $A^{\mathbf{N}}$, endowed with the product topology of the discrete topology on each copy of A, to an infinite word $w = au\varphi(u)\varphi^2(u)\varphi^3(u)\dots$. 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* (respectively *uniformly morphic*) if there exist a sequence u generated by a morphism (respectively a uniform morphism) defined over an alphabet B and a morphism (respectively a uniform morphism) φ from B to A such that $w = \varphi(u)$.

2.2. **Finite Automata and Automatic Sequences.** A formal definition of a finite automaton is given, for instance, in [15, Section 4.1], [46, Section 1.3.2], [12, Section 3.3] and [5, Section 3]. We do not give the exact definition but we propose a number of examples in Section 2.3. It suffices to say that a finite automaton consists of the following elements:

- the *input alphabet*, which is usually the set of $g \geqslant 2$ digits $\{0, 1, \dots, g-1\}$;
- the *set of states* $\mathcal{Q}$, usually a finite set of 2 or more elements, with one element, called the *initial state* and denoted by i, singled out; the elements of $\mathcal{Q}$ will be denoted by letters $\{i, a, b, \dots\}$;
- the *transition map* $\mathcal{Q} \times \{0, 1, \dots, g-1\} \to \mathcal{Q} : (a, n) \mapsto n[a]$ which associates to every state a a new state $n[a]$ depending on the current input n;
- the *output alphabet* $\mathcal{A}$, together with the *output map* $\mathcal{Q} \to \mathcal{A}$.

We extend the transition map to a map $(a, n) \mapsto n[a]$ from $\mathcal{Q} \times \mathbf{N} \to \mathcal{Q}$ as follows: let $(a, n) \in \mathcal{Q} \times \mathbf{N}$, replace n by its g-ary expansion $n = e_k e_{k-1} \dots e_1 e_0$ and define inductively $n[a] = e_k e_{k-1} \dots e_1[b]$, where $b = e_0[a]$ is the image of (a, e_0) under the transition map. Hence an automaton produces an infinite sequence

$$0[i], 1[i], 2[i], 3[i], \dots, n[i], \dots$$

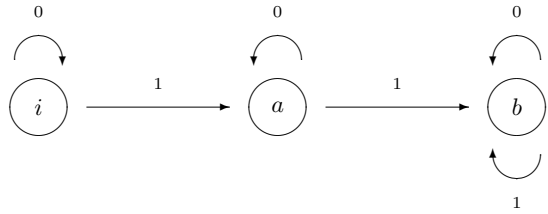
of elements of $\mathcal{Q}$. We take the images of the elements of this sequence under the transition map, and this gives rise to the *output sequence*. Such a sequence is called *g-automatic*. In other words, for an integer $g \geqslant 2$, an infinite sequence $(a_n)_{n\geqslant 0}$ of elements of $\{0, 1, \dots, g-1\}$ is said to be *g-automatic* if a_n is a finite-state function of the representation of n in base g.

For instance let us explain why the characteristic function of the sequence $2^0 = 1$, $2^1 = 2$, $2^2 = 4$, … of powers of 2 is 2-automatic. Take $g = 2$, $\mathcal{Q} = \{i, a, b\}$, and define

the transition map by

$$0[i] = i, \quad 0[a] = a, \quad 0[b] = b,$$
$$1[i] = a, \quad 1[a] = b, \quad 1[b] = b.$$

A convenient notation is given by the following diagram.



One easily checks, on writing the sequence of positive integers in base 2, that

$$0[i] = i, \quad 1[i] = a, \quad 10[i] = a, \quad 11[i] = b,$$
$$100[i] = a, \quad 101[i] = b, \quad 110[i] = b, \quad \ldots.$$

Next define $f(i) = 0$, $f(a) = 1$ and $f(b) = 0$. The output sequence

$$a_0a_1a_2\ldots = 01101000100000001000\ldots$$

is given by

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

According to Cobham [39], automatic sequences have complexity $p(m) = O(m)$; see also [15, Section 10.3]. Also, automatic sequences are the same as *uniform morphic sequences*; see for instance [15, Section 6.3] and [8, Theorem 4.1].

Automatic sequences are between periodicity and chaos. They occur in connection with harmonic analysis, ergodic theory, fractals, Feigenbaum trees, quasi-crystals, transition phases in statistical mechanics, and others; see [14, 34, 12] and [15, chapter 17].

2.3. **Examples**

2.3.1. *The Fibonacci Word* Consider the alphabet $A = \{a, b\}$. Start with $f_1 = b$, $f_2 = a$ and define $f_n = f_{n-1}f_{n-2}$. Then

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

There is a unique word

$$w = abaababaabaababaababaabaababaabaab\ldots$$

in which f_n is the prefix of length F_n, the Fibonacci number of index n, for $n \geqslant 2$. This is the *Fibonacci word*; it is generated by a binary recurrent morphism [15, Section 7.1] and 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}.

Let us check that the Fibonacci word is not periodic. Indeed, the word f_n has length F_n, and 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. On the other hand, for a periodic word the proportion of each letter is a rational number.

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

Proposition 2.1. *The Fibonacci word is Sturmian.*

We write the factors of length $1, 2, 3, 4, 5, \ldots$ of the Fibonacci word in columns as shown.

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

The k-th column contains $k+1$ elements: one of them has two right extensions to a factor of length $k+1$, and all remaining k factors of length k have a single extension. This is easily seen to be a characterization of Sturmian words.

Concerning the sequence

$$(v_n)_{n\geqslant 0} = (0,1,0,0,1,0,1,0,0,1,0,0,1,0,\ldots), \tag{5}$$

derived from the Fibonacci word on the alphabet $\{0,1\}$, a result of Danilov [41] in 1972 states that for all integers $g \geqslant 2$, the number

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

is transcendental; see also [11, Theorem 4.2].

Proposition 2.2. *The Fibonacci word is not automatic.*

Proposition 2.2 follows from a result of Cobham [39] which shows that 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 Y and a pair of adults by A. From one year to the next, the young pair become adult, which we write as $Y \to A$, while the adult pair stay alive and produce a young pair, which we write as $A \to AY$. This gives rise to the dynamical system

$$Y \to A \to AY \to AYA \to AYA\,AY \to AYA\,AYAYA \to \ldots,$$

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

$$A, Y, A, A, Y, A, Y, A, A, Y, A, A, Y, A, \ldots.$$

Replacing Y by 1 and A by 0 produces the sequence (5) considered by Danilov.

Any integer $n \geqslant 2$ has a unique representation as the sum of distinct Fibonacci numbers F_m, $m \geqslant 2$, 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 A or Y. If the smallest index in the decomposition of n is even, then $R_n = A$. If it is odd, then $R_n = Y$. For instance, for $51 = F_9 + F_7 + F_4 + F_2$, the smallest index, namely 2, is even, so $R_{51} = A$.

Recall that $\Phi = (1 + \sqrt{5})/2$ denotes the golden number. The sequence of indices n for which $R_n = A$ is

$$\lfloor \Phi \rfloor = 1, \quad \lfloor 2\Phi \rfloor = 3, \quad \lfloor 3\Phi \rfloor = 4, \quad \lfloor 4\Phi \rfloor = 6, \quad \ldots,$$

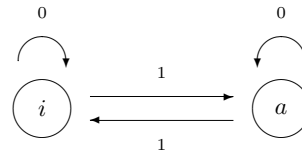
while the sequence of indices n for which $R_n = Y$ is

$$\lfloor \Phi^2 \rfloor = 2, \quad \lfloor 2\Phi^2 \rfloor = 5, \quad \lfloor 3\Phi^2 \rfloor = 7, \quad \lfloor 4\Phi^2 \rfloor = 8, \quad \ldots.$$

For instance, $32\Phi = 51.77\ldots$, so $\lfloor 32\Phi \rfloor = 51$ and $R_{51} = A$.

This sequence $(R_n)_{n \geqslant 1}$ of rabbits is an example of a *Beatty sequence*.

2.3.2. *The Prouhet–Thue–Morse word abbabaabbaababbab*$\ldots$. The finite automaton



with $f(i) = 0$ and $f(a) = 1$ produces the sequence $a_0 a_1 a_2 \ldots$, where $a_n = f(n[i])$. For instance, with $n = 9$, since 9 is 1001 in binary notation,

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

and $f(i) = 0$, we have $a_9 = 0$.

This is the *Prouhet–Thue–Morse sequence*

$$01101001100101101\ldots,$$

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, and is 0 otherwise; see [15, Section 1.6].

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

$$w = abbabaabbaababbab\ldots.$$

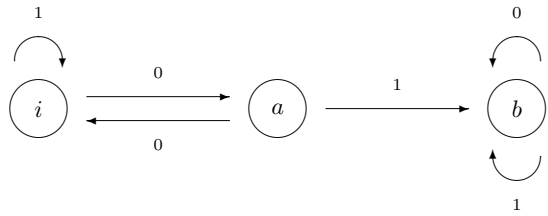
This word is generated by a binary recurrent morphism; see [15, Section 6.2]. It is the fixed point of the morphism $a \mapsto ab$, $b \mapsto ba$.

An interesting observation, due to Thue in 1906, is that if w is a finite word and a a letter for which 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 sequence, no three consecutive identical blocks such as 000 or 111 or 010101 or 101010 or 001001001 can occur.

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

$$11011001010010011001 0\ldots.$$

This sequence is automatic, associated with the automaton



with $f(i) = 1$, $f(a) = 0$ and $f(b) = 0$; see [15, Example 5.1.7].

2.3.4. *Powers of* 2. As we have seen, the binary number

$$\xi := \sum_{n\geqslant 0} 2^{-2^n} = 0.11010001000000010 00\ldots = 0.a_1a_2a_3\ldots$$

is 2-automatic. The associated infinite word

$$\mathbf{v} = v_1v_2v_3\ldots = bbabaaabaaaaaaabaaa\ldots,$$

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 given below.

m	1	2	3	4	5	6	...
$p(m)$	2	4	6	7	9	11	...

2.3.5. *The Rudin–Shapiro Word.* For $n \geqslant 0$, let $r_n \in \{a, b\}$ satisfy $r_n = a$ (respectively $r_n = b$) 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 beginning with 1 and let φ be the morphism defined from B^* to $\{a, b\}^*$ by $\varphi(1) = aa$, $\varphi(2) = ab$, $\varphi(3) = ba$ and $\varphi(4) = bb$. Then the *Rudin–Shapiro word* is $\varphi(\mathbf{u})$, hence it is morphic.

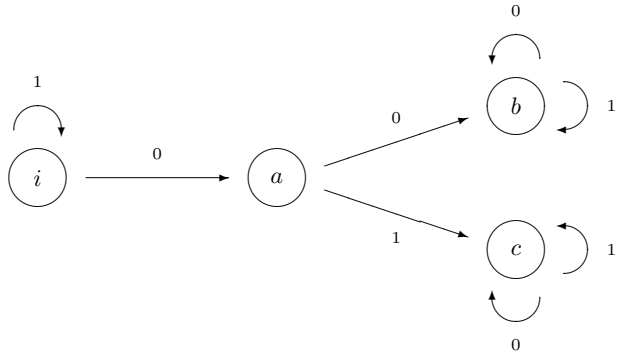
2.3.6. *Paper Folding.* Folding a strip of paper always in the same direction, and then opening it up, yields a sequence of folds which can be encoded using the digits 0 and 1. The resulting sequence $(u_n)_{n\geqslant 0}$, given by

$$1101100111001001\ldots,$$

satisfies

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

and is produced by the automaton



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

An equivalent definition for this sequence is given as follows: the sequence $a_n=u_{n+1}$, $n\geqslant 1$, is defined recursively by $a_n=1$ if n is a power of 2, and

$$a_{2^k+a}=1-a_{2^k-a}, \quad 1\leqslant a<2^k;$$

see [15, Example 5.1.6].

For a connection between the paper folding sequence and the Prouhet–Thue–Morse sequence, see [18].

2.4. **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 Ferenczi and Mauduit [42] in 1997. The point is that such sequences contain sequences of digits which bear similarities, and yields the existence of very sharp rational approximations which do not exist for algebraic numbers.

It follows from their work that the complexity of the g-ary expansion of every irrational algebraic number satisfies

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

see [10]. The main tool for the proof is a p-adic version of the Thue–Siegel–Roth theorem due to Ridout in 1957; see Theorem 4.3 below as well as [3].

Several papers have been devoted to the study of the complexity of the g-ary expansions of real algebraic numbers, in particular by Allouche and Zamboni in 1998, Risley and Zamboni in 2000, and Adamczewski and Cassaigne in 2003. For a survey, see [3]. The main recent result is the following [5].

Theorem 2.3 (Adamczewski and Bugeaud, 2007). *The complexity $p(m)$ of the g-ary expansion of a real irrational algebraic number satisfies*

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

In 1968,[1] Cobham [38] claimed that automatic irrational numbers are transcendental. This follows from Theorem 2.3, since automatic numbers have a complexity $O(m)$.

Corollary 2.4 (Conjecture of Cobham). *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 2.3 is a new, combinatorial transcendence criterion obtained by Adamczewski, Bugeaud and Luca [7] as an application of Schmidt's subspace theorem; see Theorem 4.4 and [60].

In 1979, Christol [36] proved that when p is a prime and

$$f(X) = \sum_{k\geqslant 1} u_k X^k \in \mathbf{F}_p[[X]]$$

a power series with coefficients in the finite field $\mathbf{F}_p$, then f is algebraic if and only if the sequence of coefficients $(u_k)_{k\geqslant 1}$ can be produced by a p-automaton.

According to Cobham, a sequence can be simultaneously k-automatic and ℓ-automatic with k and ℓ multiplicatively independent if and only if it is ultimately periodic. Christol, Kamae, Mendès France and Rauzy [37] have shown that a power series with coefficients in a finite set can be algebraic over two different finite fields if and only if it is rational.

In this context Theorem 2.3 implies the following statement.

Corollary 2.5. *Let p be a prime number, $g \geqslant p$ an integer and $(u_k)_{k\geqslant 1}$ a sequence of integers in the range $\{0,\dots,p-1\}$. The formal power series*

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

and the real number

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

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

As an example, taken from [11, Section 6] and [12, Section 2.4]), consider the Prouhet–Thue–Morse sequence $(a_n)_{n\geqslant 0}$. The series

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

is algebraic over $\mathbf{F}_2(X)$, as it is a root of $(1+X)^3F^2 + (1+X)^2F + X = 0$. Since F is not a rational function, Corollary 2.5 gives another proof of Mahler's transcendence result on the number

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

[1] The author is grateful to Boris Adamczewski for this reference.

The following result related to Theorem 2.3 has been obtained subsequently by Bugeaud and Evertse [32] by means of a refinement of the so-called Cugiani–Mahler theorem. The main tool is again a quantitative version of Schmidt's subspace theorem [60].

Theorem 2.6 (Bugeaud and Evertse, 2007). *Let $b \geqslant 2$ be an integer and ξ an irrational algebraic number with $0 < \xi < 1$. Then for any real number $\eta < 1/11$, the complexity $p(m)$ of the b-ary expansion of ξ satisfies*

$$\limsup_{m\to+\infty} \frac{p(m)}{m(\log m)^\eta} = +\infty.$$

Further developments of the method of [7, 5] have recently been achieved by Adamczewski and Rampersad [9] who have shown that the binary expansion of an algebraic number contains infinitely many occurrences of 7/3-powers. Here, for a positive real number w, a *w-th power* is a word of the form $W^{\lfloor w\rfloor}W'$, where W is a word of length $|W|$, say, and W' is the prefix of W of length $\lceil (w-\lfloor w\rfloor)|W|\rceil$. For instance with $w = 7/3$ and $W = 011010$ we have $|W| = 6$, $\lfloor w\rfloor = 2$ and $\lceil (w-\lfloor w\rfloor)|W|\rceil = 2$, so 011010 011010 01 is a 7/3-th power.

Adamczewski and Rampersad deduce from their result that the binary expansion of an algebraic number contains infinitely many occurrences of overlaps, where an overlap is a pattern of the form $xXxXx$, where x is a letter and X a word.

3. Analytic Methods

3.1. **Mahler's Method.** Mahler [48] initiated a new transcendence method in 1929 for studying values of functions satisfying certain functional equations. His work was not widely known and in 1969 he published [49] which was the source of a revival of this method; see [47, 52].

A first example [52, Theorem 1.1.2] is the function

$$f(z) = \sum_{n\geqslant 0} z^{-d^n}, \quad d \geqslant 2,$$

which satisfies the functional equation $f(z^d) + z = f(z)$ for $|z| < 1$. Mahler proved that it takes transcendental values at algebraic points in the domain $0 < |z| < 1$. A special case is (3).

Mahler also proved in 1929 that the so-called Prouhet–Thue–Morse–Mahler number in base $g \geqslant 2$, given by

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

where $(a_n)_{n\geqslant 0}$ is the Prouhet–Thue–Morse sequence, is transcendental; see [52] and [15, Section 13.4]. The idea of proof is as follows; see [15, Section 13.4] and [52, Example 1.3.1], where the complete proof is given. Consider the function

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

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

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

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

Another application [20] of Mahler's method is the transcendence of the number (4) whose digits in a given base g are 1 at the Fibonacci indices $1, 2, 3, 5, 8, \ldots$, and 0 elsewhere.

After Cobham [39], Loxton and van der Poorten also tried[2] in 1982 and 1988 to use Mahler's method to prove Cobham's conjecture, now Corollary 2.4 of Theorem 2.3 of Adamczewski and Bugeaud. Becker [25] pointed out in 1994 that Mahler's method yields only a weaker result so far, that for any given non-eventually periodic automatic sequence $\mathbf{u} = (u_1, u_2, u_3 \ldots)$, the real number

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

is transcendental, provided that the integer g is sufficiently large in terms of $\mathbf{u}$; see also [5]. It is yet a challenge to extend Mahler's method in order to prove Cobham's conjecture.

There is a further very interesting development of Mahler's method which we only allude to here without entering the subject. It is due to Denis and deals with transcendence problems in finite characteristic; see Pellarin's report [53] in the Bourbaki Seminar.

3.2. **Nesterenko's Theorem and Consequences.** The transcendence of the Liouville-Fredholm number (7) below was studied only ten years ago by Bertrand in 1997, and Duverney, Nishioka as well as Nishioka and Shiokawa in 1998, as a consequence of the results due to Nesterenko's work in 1996 on the transcendence of values of theta series at rational points involving modular functions; see [51, Chapter 3, Section 1.3]. It follows from these works that for algebraic α in the domain $0 < |\alpha| < 1$, the three numbers

$$\sum_{n\geqslant 0}\alpha^{-n^2}, \quad \sum_{n\geqslant 0} n^2\alpha^{-n^2}, \quad \sum_{n\geqslant 0} n^4\alpha^{-n^2}$$

are algebraically independent.

A related example from [5] is the number

$$\eta = \sum_{k\geqslant 1} u_k 3^{-k}$$

which is associated to the word

$$\mathbf{u} = 012^1 12^2 12^3 12^4 12^5 12^6 12^7 \ldots,$$

where, for instance, 2^4 denotes 2222, generated by the non-recurrent morphism $0 \mapsto 012$, $1 \mapsto 12$, $2 \mapsto 2$. Using

$$\eta = 1 - \sum_{n\geqslant 1} 3^{-n(n+1)/2},$$

[2] See [47] and [15, Section 13.10].

one deduces from Nesterenko's result that it is transcendental. For this number the growth of complexity $p(m)$ is quadratic in m, so Theorem 2.3 does not apply.

4. Diophantine Approximation

4.1. **Liouville, Thue, Siegel, Roth, Ridout, Schmidt.** Diophantine approximation theory yields information on the arithmetic nature of numbers of the form

$$\sum_{n\geqslant 0} g^{-u_n}. \tag{6}$$

For instance, if the sequence satisfies

$$u_{n+1} - u_n \to +\infty, \quad n \to +\infty,$$

then the number given by (6) is irrational. This follows from the fact that a real number is rational if and only if its g-ary expansion is ultimately periodic. It also follows from Diophantine analysis. As a special case take $u_n = n^2$. The irrationality of the Liouville–Fredholm number

$$\theta = \sum_{n\geqslant 0} g^{-n^2} \tag{7}$$

follows.

Here is the very simple proof, which goes back to Liouville in 1851; see [51, Chapter 3, Section 1.3]. Let $\epsilon > 0$. Let N be a sufficiently large integer. Set

$$q = a^{N^2} \quad \text{and} \quad p = \sum_{n=0}^{N} a^{N^2-n^2}.$$

Then

$$0 < q\theta - p = \sum_{n\geqslant N+1} \frac{1}{a^{n^2-N^2}} = \sum_{k\geqslant 1} \frac{1}{a^{2Nk+k^2}} < \frac{\theta - 1}{a^{2N}}.$$

For N sufficiently large the right-hand side is less than ϵ and the irrationality of θ is proved.

That θ is not a quadratic irrationality follows from the work of Bailey and Crandall [22]. See Section 3.2 for the transcendence of this number θ.

The first transcendence statement on numbers given by a series (6), assuming that the sequence $(u_n)_{n\geqslant 0}$ is increasing and grows sufficiently fast, goes back to Liouville in 1844.

Theorem 4.1 (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 (6), provided that the increasing sequence $(u_n)_{n\geqslant 0}$ satisfies

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

For instance, $u_n = n!$ satisfies this condition, so the number

$$\sum_{n\geqslant 0} g^{-n!}$$

is transcendental.

Theorem 4.2 (Thue, Siegel, Roth [57, 58], 1955). *For any real algebraic number α and any $\epsilon > 0$, the set of $p/q \in \mathbf{Q}$ with $|\alpha - p/q| < q^{-2-\epsilon}$ is finite.*

From Theorem 4.2, one deduces the transcendence of the series (6) under the weaker hypothesis

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

The sequence $u_n = \lfloor \kappa^n \rfloor$ satisfies this condition as soon as $\kappa > 2$. For example, the transcendence of the number

$$\sum_{n\geqslant 0} g^{-3^n}$$

follows from Theorem 4.2.

A stronger result follows from Theorem 4.3 below due to Ridout (see, for instance, [15, Section 13.5]), using the fact that the denominators g^{u_n} are powers of g. The condition

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

is sufficient to imply the transcendence of the sum of the series (6); see [1]. An example is the transcendence of

$$\sum_{n\geqslant 0} g^{-2^n};$$

see also (3) above.

Theorem 4.3 (Ridout, 1957). *For any $g \geqslant 2$, any real algebraic number α and any $\epsilon > 0$, the set of $p/q \in \mathbf{Q}$ with $q = g^k$ 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 [60], established in 1972, together with its p-adic extension by Schlickewei in 1976. We state only a simplified version; see Bilu's Bourbaki lecture [27] for references and further recent achievements based on this fundamental result.

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

Theorem 4.4 (Schmidt's subspace theorem). *Let $m \geqslant 2$ be an 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 $\mathbf{x} = (x_0, \dots, x_{m-1}) \in \mathbf{Z}^m$ for which*

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

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

Theorem 4.2 due to Thue-Siegel-Roth follows from Theorem 4.4 if one takes

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

Theorem 4.3 due to Ridout is also a consequence of Theorem 4.4 if one takes

$$S = \{\infty\} \cup \{\ell : \ell \text{ prime and } \ell \mid g\}$$

and

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

Since there is no loss of generality to assume that g is squarefree, so that

$$g = \prod_{\ell | g} \ell,$$

for $(x_0, x_1) = (q, p)$ with $q = g^k$, we have

$$|L_{0,\infty}(x_0, x_1)|_\infty = q, \quad |L_{1,\infty}(x_0, x_1)|_\infty = |q\alpha - p|,$$
$$\prod_{\ell | g} |L_{0,\ell}(x_0, x_1)|_\ell = q^{-1}, \quad \prod_{\ell | g} |L_{1,\ell}(x_0, x_1)|_\ell = |p|_\ell \leqslant 1.$$

Further applications of the Subspace theorem to transcendence questions have been obtained by Corvaja and Zannier [40].

4.2. **Irrationality and Transcendence Measures.** The previous results can be made effective in order to reach irrationality measures or transcendence measures for automatic numbers.

In 2006, Adamczewski and Cassaigne [8] solved a conjecture of Shallit in 1999 by proving that the sequence of g-ary digits of 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 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 a real irrational algebraic number has irrationality exponent 2 by Theorem 4.2, as do almost all real numbers.

An explicit upper bound for the irrationality exponent for automatic irrational numbers is given by [8, Theorem 2.2]. For the Prouhet–Thue–Morse–Mahler numbers for instance, the exponent of irrationality is at most 5. However, there is no uniform upper bound for such exponents, as pointed out to the author by 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, Adamczewski and Bugeaud [6] have obtained transcendence measures for automatic numbers. They show that automatic irrational numbers are either S- or T-numbers in Mahler's classification of transcendental numbers; see [30]. This is a partial answer to a conjecture of Becker [8] which states that all automatic irrational numbers are S-numbers.

5. Continued Fractions

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

5.1. **Complexity of the Continued Fraction Expansion of an Algebraic Number.** In 1949, Khinchin asked the following question: Are the partial quotients of the continued fraction expansion of a real non-quadratic irrational algebraic number bounded? So far no example is known. It is not yet ruled out that all these partial quotients are bounded for all such x, or that they are unbounded for all such x. The common expectation seems to be that they are never bounded. The situation in finite characteristic is quite different. In 1976, Baum and Sweet [23] constructed a formal series which is cubic over the field $\mathbf{F}_2(X)$, the continued fractions of which have partial quotients of bounded degree.

We give here only a very short historical account. The very first transcendence results for numbers given by their continued fraction expansions are due to Liouville in 1844. This topic was extensively developed by Maillet in 1906 and later by Perron in 1929. In 1955, the Thue–Siegel–Roth theorem enabled Davenport and Roth to obtain deeper results. Further investigations are due to Baker in 1962 and 1964. The approximation results on real numbers by quadratic numbers due to Schmidt in 1967 are a main tool for the next steps by Davison in 1989, and by Queffélec [55] who established in 1998 the transcendence of the Prouhet–Thue–Morse continued fraction; see also [15, Section 13.7]. There are further papers by Liardet and Stambul in 2000 and by Baxa in 2004. In 2001, Allouche, Davison, Queffélec and Zamboni [13] proved the transcendence of Sturmian continued fractions, namely continued fractions whose sequence of partial quotients is Sturmian. The transcendence of the Rudin–Shapiro and of the Baum–Sweet continued fractions was proved in 2005 by Adamczewski, Bugeaud and Davison. In 2005, Adamczewski and Bugeaud [2] showed that the continued fraction expansion of an algebraic number of degree at least three cannot be generated by a binary morphism.

5.2. **The Fibonacci Continued Fraction.** The Fibonacci word discussed in Section 2.3.1 enabled Roy [59] to construct transcendental real numbers ξ for which ξ and ξ^2 are surprisingly well simultaneously and uniformly approximated by rational numbers.

Recall once more that Φ denotes the golden number, so that

$$\Phi^{-1} = \Phi - 1 = \frac{\sqrt{5}-1}{2} = 0.618\ldots.$$

Theorem 5.1 (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 respectively to obtain*

$$[0; A, B, A, A, B, A, B, A, A, B, A, A, B, A, B, A, A, B, A, B, A,$$
$$A, B, A, A, B, A, B, A, A, B, A, A, B \ldots].$$

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

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

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

The arguments of Roy provide simplified proofs of Queffélec's results on the transcendence of the Thue–Morse and Fibonacci continued fractions.

Using Theorem 5.1, together with ideas of Davenport and Schmidt involving a transference theorem concerning Mahler's convex bodies, 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 obtained by Bugeaud and Laurent, Fischler, Roy, and more recently also by Adamczewski and Bugeaud [6].

For further references on Diophantine approximation, we refer the reader to Bugeaud's recent book [30].

6. Open problems

Many problems related to the present topic are open. To trace their original sources would require more thorough bibliographical investigation. While the origins of these questions are most often much older, we give references to recent papers where they are quoted.

Among the open questions raised in [2] is the following:

- Does there exist an algebraic number of degree at least 3 whose continued fraction expansion is generated by a morphism?

In the same vein the next question is open:

- Do there exist an integer $g \geqslant 3$ and an algebraic number of degree at least 3 whose expansion in base g is generated by a morphism?

Two open questions, for which positive answers are expected, are proposed in [9]:

- Is it true that the binary expansion of every algebraic number contains arbitrarily large squares?
- Is it true that the binary expansion of every algebraic number contains arbitrarily large palindromes?

Recall that a square is nothing other than a 2-power, namely a pattern XX where X is a word, while a palindrome is a word $W = w_1 w_2 \dots w_r$ which is invariant under reversal: if $\overline{W}$ denotes the word $w_r \dots w_2 w_1$, then $W = \overline{W}$. Hence a palindrome is either of the form $X\overline{X}$ or of the form $Xx\overline{X}$ where X is a word and x a letter.

Other problems are suggested by Rivoal [56]:

- Let $g \geqslant 2$ be an integer. Give an explicit example of a real number $x > 0$ which is simply normal in base g and for which $1/x$ is not simply normal in base g.
- Let $g \geqslant 2$ be an integer. Give an explicit example of a real number $x > 0$ which is normal in base g and for which $1/x$ is not normal in base g.
- Give an explicit example of a real number $x > 0$ which is normal and for which $1/x$ is not normal.

Remark. In [45], there is a construction of an automatic number, the inverse of which is not automatic. This answers by anticipation [15, Section 13.9, Problem 2].

From the open problems in [15, Section 13.9], we select the following two:

- Show that the number $\log 2$ is not 2-automatic; see (1).
- Show that the number π is not 2-automatic; see (2).

We conclude this chapter with one last open problem, attributed to Mahler; see for instance [4]:

- Let $(e_n)_{n\geqslant 1}$ be an infinite sequence over $\{0,1\}$ that is not ultimately periodic. Is it true that at least one of the two numbers

$$\sum_{n\geqslant 1} e_n 2^{-n} \quad \text{and} \quad \sum_{n\geqslant 1} e_n 3^{-n}$$

  is transcendental?

From Conjecture 1.1 with $g = 3$ and $a = 2$, it follows that the second number should be always transcendental.

Acknowledgments. The author is grateful to Boris Adamczewski, Jean–Paul Allouche, Yann Bugeaud, Stéphane Fischler, Damien Roy, Tanguy Rivoal and Paul Voutier for their useful comments on a preliminary version of this paper. This topic was the subject of a lecture by the author at a conference in honour of the 65th birthday of Professor Iekata Shiokawa, held at Keio University, Yokohama, Japan, on 7–10 March 2006. The author is also grateful to Takao Komatsu for his invitation to participate at this conference. A tremendous job was done by William Chen who edited and improved substantially the submitted paper before publication in this volume.

References

[1] B. Adamczewski. Transcendance "à la Liouville" de certains nombres réels. *C. R. Math. Acad. Sci. Paris*, 338 (2004), 511–514.
[2] B. Adamczewski, Y. Bugeaud. On the complexity of algebraic numbers II: Continued fractions. *Acta Math.*, 195 (2005), 1–20.
[3] B. Adamczewski, Y. Bugeaud. On the decimal expansion of algebraic numbers. *Fiz. Mat. Fak. Moksl. Semin. Darb.*, 8 (2005), 5–13.
[4] B. Adamczewski, Y. Bugeaud. Real and p-adic expansions involving symmetric patterns. *Int. Math. Res. Not.*, (2006), article ID 75968, 17 pages.
[5] B. Adamczewski, Y. Bugeaud. On the complexity of algebraic numbers I: Expansion in integer bases. *Ann. of Math.*, 165 (2007), 547–565.
[6] B. Adamczewski, Y. Bugeaud. Mesures de transcendance et aspects quantitatifs de la méthode de Thue-Siegel-Roth-Schmidt (preprint).
[7] B. Adamczewski, Y. Bugeaud, F. Luca. Sur la complexité des nombres algébriques. *C. R. Math. Acad. Sci. Paris*, 339 (2004), 11–14.
[8] B. Adamczewski, J. Cassaigne. Diophantine properties of real numbers generated by finite automata. *Compos. Math.*, 142 (2006), 1351–1372.
[9] B. Adamczewski, N. Rampersad. On patterns occurring in binary algebraic numbers. *Proc. Amer. Math. Soc.* (to appear).
[10] J.-P. Allouche. Nouveaux résultats de transcendance de réels à développement non aléatoire. *Gaz. Math.*, 84 (2000), 19–34.
[11] J.-P. Allouche. Automates et algébricités. *J. Théor. Nombres Bordeaux*, 17 (2005), 1–11.
[12] J.-P. Allouche. Suites automatiques et séries formelles algébriques. *Mathématiques pour l'Ingénieur, Techniques de l'Ingénieur*, Dossier AF175 (2005), AF175-1–AF175-8.
[13] J.-P. Allouche, J.L. Davison, M. Queffélec, L.Q. Zamboni. Transcendence of Sturmian or morphic continued fractions. *J. Number Theory*, 91 (2001), 39–66.
[14] J.-P. Allouche, M. Mignotte. Arithmétique et automates. *Images des Mathématiques 1988*, pp. 5–9 (Courrier du CNRS, supplément au numéro 69, Centre National de la Recherche Scientifique, 1988).
[15] J.-P. Allouche, J. Shallit. *Automatic sequences: Theory, Applications, Generalizations* (Cambridge University Press, 2003).

[16] P. Arnoux, C. Mauduit, I. Shiokawa, J.-I. Tamura. Complexity of sequences defined by billiard in the cube. *Bull. Soc. Math. France*, 122 (1994), 1–12.

[17] P. Arnoux, C. Mauduit, I. Shiokawa, J.-I. Tamura. Rauzy's conjecture on billiards in the cube. *Tokyo J. Math.*, 17 (1994), 211–218.

[18] R. Bacher. La suite de Thue-Morse et la catégorie Rec. *C. R. Math. Acad. Sci. Paris*, 342 (2006), 161–164.

[19] D.H. Bailey, J.M. Borwein. Experimental mathematics: recent developments and future outlook. *Mathematics Unlimited – 2001 and Beyond*, (B. Engquist, W. Schmid, eds.) pp. 51–66 (Springer-Verlag, 2001).

[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), 487–518.

[21] D.H. Bailey, P. Borwein, S. Plouffe. On the rapid computation of various polylogarithmic constants. *Math. Comp.*, 66 (1997), 903–913.

[22] D.H. Bailey, R.E. Crandall. On the random character of fundamental constant expansions. *Experiment. Math.*, 10 (2001), 175–190.

[23] L.E. Baum, M.M. Sweet. Continued fractions of algebraic power series in characteristic 2. *Ann. of Math.*, 103 (1976), 593–610.

[24] V. Becher, S. Figueira. An example of a computable absolutely normal number. *Theoret. Comput. Sci.*, 270 (2002), 947–958.

[25] P.-G. Becker. k-regular power series and Mahler-type functional equations. *J. Number Theory*, 49 (1994), 269–286.

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

[27] Y. Bilu. The many faces of the subspace theorem [after Adamczewski, Bugeaud, Corvaja, Zannier, ...]. *Séminaire Bourbaki, Exposé 967, 59é année* (2006-2007).

[28] É. Borel. Les probabilités dénombrables et leurs applications arithmétiques. *Rend. Circ. Mat. Palermo*, 27 (1909), 247–271.

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

[30] Y. Bugeaud. *Approximation by Algebraic Numbers* (Cambridge Tracts in Mathematics 160, Cambridge University Press, 2004).

[31] Y. Bugeaud. On the b-ary expansion of an algebraic number. *Rend. Sem. Math. Univ. Padova* (to appear).

[32] Y. Bugeaud, J.H. Evertse. On two notions of complexity of algebraic numbers (manuscript). `http://arxiv.org/abs/0709.1560`

[33] C.S. Calude. *Information and Randomness: An Algorithmic Perspective*, 2nd edition (Texts in Theoretical Computer Science – an EATCS Series, Springer-Verlag, 2002).

[34] R. Cerf. Le modèle d'Ising et la coexistence des phases. *Images des Mathématiques 2004*, pp. 47–51 (Courrier du CNRS, Centre National de la Recherche Scientifique, 2004).

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

[36] G. Christol. Ensembles presque périodiques k-reconnaissables. *Theoret. Comput. Sci.*, 9 (1979), 141–145.

[37] G. Christol, T. Kamae, M. Mendès France, G. Rauzy. Suites algébriques, automates et substitutions. *Bull. Soc. Math. France*, 108 (1980), 401–419.

[38] A. Cobham. On the Hatmanis-Stearns problem for a class of tag machine. *IEEE Conference Record of 1968 Ninth Annual Symposium on Switching and Automata Theory, Schenectady, New York, 1968*, pp. 51–60 (IEEE Computer Society Press, 1968).

[39] A. Cobham. Uniform tag sequences. *Theory Comput. Syst.*, 6 (1972), 164–192.

[40] P. Corvaja, U. Zannier. Some new applications of the subspace theorem. *Compos. Math.*, 131 (2002), 319–340.

[41] L.V. Danilov. Certain classes of transcendental numbers. *Mat. Zametki*, 12 (1972), 149–154.

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

[43] M. Kontsevich, D. Zagier. Periods. *Mathematics Unlimited – 2001 and Beyond*, (B. Engquist, W. Schmid, eds.) pp. 771–808 (Springer-Verlag, 2001).
[44] J.C. Lagarias. On the normality of arithmetical constants. *Experiment. Math.*, 10 (2001), 355–368.
[45] S. Lehr, J. Shallit, J. Tromp. On the vector space of the automatic reals. *Theoret. Comput. Sci.*, 163 (1996), 193–210.
[46] M. Lothaire. *Algebraic Combinatorics on Words* (Encyclopedia of Mathematics and its Applications 90, Cambridge University Press, 2002).
[47] J.H. Loxton, A.J. van der Poorten. Arithmetic properties of certain functions in several variables III. *Bull. Austral. Math. Soc.*, 16 (1977), 15–47.
[48] K. Mahler. Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen. *Math. Ann.*, 101 (1929), 342–366; 103 (1930), 532.
[49] K. Mahler. Remarks on a paper by W. Schwarz. *J. Number Theory*, 1 (1969), 512–521.
[50] K. Mahler. Arithmetical properties of the digits of the multiples of an irrational number. *Bull. Austral. Math. Soc.*, 8 (1973), 191–203.
[51] Yu.V. Nesterenko, P. Philippon, eds. *Introduction to Algebraic Independence Theory* (with contributions from F. Amoroso, D. Bertrand, W.D. Brownawell, G. Diaz, M. Laurent, Yu.V. Nesterenko, K. Nishioka, P. Philippon, G. Rémond, D. Roy and M. Waldschmidt) (Lecture Notes in Mathematics 1752, Springer-Verlag, 2001).
[52] K. Nishioka. *Mahler Functions and Transcendence* (Lecture Notes in Mathematics 1631, Springer-Verlag, 1996).
[53] F. Pellarin. Aspects de l'indépendance algébrique en caractéristique non nulle [d'après Anderson, Brownawell, Denis, Papanikolas, Thakur, Yu, ...]. *Séminaire Bourbaki, Exposé 973, 59é année* (2006-2007).
[54] S.S. Pillai. On normal numbers. *Proc. Indian Acad. Sci. A*, 10 (1939), 13–15; 12 (1940), 179–184.
[55] M. Queffélec. Transcendance des fractions continues de Thue-Morse. *J. Number Theory*, 73 (1998), 201–211.
[56] T. Rivoal. On the bits counting function of real numbers. *J. Aust. Math. Soc.* (to appear).
`http://www-fourier.ujf-grenoble.fr/~rivoal/articles/binreal.pdf`
[57] K.F. Roth. Rational approximations to algebraic numbers. *Mathematika*, 2 (1955), 1–20; corrigendum, 168.
[58] K.F. Roth. Rational approximations to algebraic numbers. *Proceedings of the International Congress of Mathematicians 1958* (J.A. Todd, ed.), pp. 203–210 (Cambridge University Press, 1960).
[59] D. Roy. Approximation simultanée d'un nombre et de son carré. *C. R. Math. Acad. Sci. Paris*, 336 (2003), 1–6.
[60] W.M. Schmidt. *Diophantine Approximation* (Lecture Notes in Mathematics 785, Springer-Verlag, 1980).
[61] T. Schneider. *Einführung in die Transzendenten Zahlen* (Grundlehren der Mathematischen Wissenschaften 81, Springer-Verlag, 1957).
[62] I. Shiokawa, J.-I. Tamura. Description of sequences defined by billiards in the cube. *Proc. Japan Acad. Ser. A Math. Sci.*, 68 (1992), 207–211.