Workshop on combinatorics on words, Montreal, March 12th, 2007

Diophantine properties of automatic real numbers

Julien Cassaigne

Institut de mathématiques de Luminy - CNRS, Marseille, France
cassaigne@iml.univ-mrs.fr

Joint work with Boris Adamczewski,
Institut Camille Jordan - CNRS, Lyon, France

Diophantine properties of automatic real numbers

- Combinatorial vs. arithmetic properties of numbers
- Borel's conjecture
- Transcendence of numbers with Sturmian expansions
- Transcendence of low-complexity numbers
- Automatic numbers
- Automatic numbers are not Liouville numbers
- Generalizations

b-adic expansion

Let $\xi \in \mathbb{R}$. For simplicity, assume $0 \leq \xi < 1$.

Fix $b \in \mathbb{N}$, $b \geq 2$. Then ξ can be written as

$$\xi = \sum_{n=1}^{\infty} x_n b^{-n}$$

where $0 \leq x_n < b$

Moreover, the sequence (x_n) is unique if we impose that $x_n \neq b - 1$ for infinitely many n. It is called the b-adic expansion of ξ.

b-adic expansions as infinite words

Let $A = \{0, 1, \ldots, b-1\}$ be the alphabet of b-adic digits.

The b-adic expansion of ξ can then be viewed as an infinite word $\mathbf{x} = x_1 x_2 x_3 \ldots \in A^\omega$.
We thus have a bijection between $[0, 1)$ and $A^\omega \setminus A^*(b-1)^\omega$.

General problem
If ξ is defined by some equation, can we say something about $\mathbf{x}$?
Conversely, if $\mathbf{x}$ is defined by some generating device, can we say something about ξ?

Yes: $\xi \in \mathbb{Q}$ iff $\mathbf{x}$ is eventually periodic. What else?

Algebraic numbers

The real number ξ is algebraic if it is the root of a polynomial with integer coefficients.

Meta-conjecture
The b-adic expansion of an algebraic irrational number has all the combinatorial properties of a random infinite word.

Conjecture [Émile Borel 1909]
Let ξ be algebraic irrational. Then it is normal in every base b, i.e., every word $w \in A^*$ occurs in $\mathbf{x}$ with frequency $b^{-|w|}$.

Borel's conjecture: discussion

Borel's conjecture seems currently out of reach. We do not even know that every digit of A occurs in $\mathbf{x}$ when $b \geq 3$.

Actually, we do not know a single example of a triplet (ξ, b, a) with ξ algebraic irrational, $b \geq 3$, and $a \in A$, for which we can prove that the digit a occurs infinitely often in the b-adic expansion of ξ, while Borel's conjecture implies that it should be true for all such triplets.

What we can do: prove that certain very particular classes of infinite words do not contain the expansion of any algebraic irrational.

Subword complexity

Let $\mathbf{u} = u_1 u_2 u_3 \ldots$ be an infinite word.

A finite word $w \in A^*$ is a factor of $\mathbf{u}$ if $w = u_k u_{k+1} \ldots u_{k+n-1}$ for some k and n.

The subword complexity of $\mathbf{u}$ is the function $p_{\mathbf{u}}$ that maps n to the number $p_{\mathbf{u}}(n)$ of factors of length n of $\mathbf{u}$.

Conjecture (weaker than Borel's, but still out of reach)
If ξ is algebraic irrational, then $p_{\mathbf{x}}(n) = b^n$ for all n.

Sturmian expansions

An infinite word $\mathbf{u}$ is Sturmian if $p_{\mathbf{u}}(n) = n + 1$ for all n.
This is the lowest possible complexity for non-periodic words.

Theorem [Ferenczi-Mauduit 1997]
If $\mathbf{x}$ is a Sturmian word, then ξ is transcendental.

Proof on a simple example, with $b = 2$: the Fibonacci word.
Let $\mathbf{x}$ be the fixed point of the substitution $\varphi : 0 \mapsto 01,\ 1 \mapsto 0$:

$$\mathbf{x} = 0100101001001010010100100101001001\ldots$$

Sketch of the proof:

- find good rational approximations of ξ;
- apply a transcendence criterion like Roth's theorem.

Roth's theorem

Theorem [Roth 1955]
Let $\xi \in \mathbb{R} \setminus \mathbb{Q}$, and $\varepsilon > 0$. If there are infinitely many rational numbers $\frac{p}{q}$ such that

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

then ξ is transcendental.

(For general Sturmian words, this is actually not enough; Ferenczi and Mauduit use a p-adic variant of this theorem, Ridout's theorem [Ridout 1957].)

Initial repetitions in the Fibonacci word

Observe that $\mathbf{x}$ starts with 01001.01001.0:

$$\mathbf{x} = 0100101001001010010100100101001001\ldots$$

Consequently, it starts with $\varphi^k(01001)\varphi^k(01001)\varphi^k(0)$ for all k.
Let $F_k = |\varphi^k(0)|$ (Fibonacci numbers), then $|\varphi^k(01001)| = F_{k+3}$.

Consider the rational number $\eta = \frac{p}{q}$ with expansion $\mathbf{y} = (\varphi^k(01001))^\omega$.
It is obviously a good approximation of ξ: $|\xi - \eta| < b^{-2F_{k+3}-F_k}$, since $\mathbf{x}$ and $\mathbf{y}$ have a common prefix of length $2F_{k+3} + F_k$.

Note that $q = b^{F_{k+3}} - 1$. As F_k/F_{k+3} tends to $\Phi^{-3} > 0.2$, for k large enough we have $|\xi - \eta| < q^{-2.2}$, so we can apply Roth's theorem.

Stammering words

Let $\kappa > 1$. The infinite word $\mathbf{x}$ is said to be κ-stammering if there exists a sequence (u_n, v_n, w_n) of triplets of words such that

- $u_n v_n w_n$ is a prefix of $\mathbf{x}$;
- w_n is a prefix of $v_n w_n$;
- $|v_n|$ is not bounded;
- $|w_n|/|v_n| \geq \kappa - 1$;
- $|u_n|/|v_n|$ is bounded.

The proof of Ferenczi and Mauduit uses the fact that Stumian words are $(2+\varepsilon)$-stammering, and then Ridout's theorem ensures that irrational numbers with a $(2+\varepsilon)$-stammering expansion are transcendental.

The theorem of Adamczewski and Bugeaud

Using a p-adic version of Schmidt's subspace theorem [Wolfgang Schmidt 1967] instead of Ridout's theorem allows a great improvement:

Theorem [Adamczewski-Bugeaud 2004]
If $\mathbf{x}$ is $(1+\varepsilon)$-stammering, then ξ is rational or transcendental.

Corollary [Adamczewski-Bugeaud 2004]
If $p_{\mathbf{x}}(n) = O(n)$, then ξ is rational or transcendental.

Theorem [Bugeaud 2007]
There exists $\gamma_b > 0$ such that, if $p_{\mathbf{x}}(n) = O(n(\log n)^{\gamma_b})$,
then ξ is rational or transcendental.

Low-complexity words are $(1+\varepsilon)$-stammering

Corollary [Adamczewski-Bugeaud 2004]
If $p_{\mathbf{x}}(n) = O(n)$, then ξ is rational or transcendental.

Proof
Assume that $p_{\mathbf{x}}(n) \leq cn$ for $n \geq 1$.
By the pigeon-hole principle, the prefix of length $p_{\mathbf{x}}(n) + n$ of $\mathbf{x}$ contains two occurrences of the same factor of length n. Let w_n be this factor, u_n the prefix up to the first occurrence of w_n, and v_n the word starting at the first occurrence of w_n (included) and ending before the second occurrence. Then the triplets (u_n, v_n, w_n) show that $\mathbf{x}$ is $(1 + 1/c)$-stammering.

Automatic numbers

Let $k \geq 2$ be an integer.
An infinite word $\mathbf{u} \in A^\omega$ is said to be k-automatic if there exists a finite automaton that takes as input the k-adic expansion of an integer n, and outputs the n-th letter of $\mathbf{u}$.

Equivalently, $\mathbf{u}$ is k-automatic iff there exists an alphabet B, two k-uniform substitutions $f : B^* \to B^*$ and $g : B^* \to A^*$ (i.e., $|f(a)| = |g(a)| = k$ for all $a \in B$) and an infinite word $\mathbf{v} \in B^\omega$ such that $\mathbf{v}$ is a fixed point of f and $\mathbf{u} = g(\mathbf{v})$.

A number ξ is said to be k-automatic in base b if its b-adic expansion is k-automatic.

Diophantine properties of automatic numbers

As k-automatic words have complexity $O(n)$, by the theorem of Adamczewski and Bugeaud all automatic numbers are either rational or transcendental.

This property was conjectured by Cobham [1968] and announced as solved by Loxton and van der Poorten [1982, 1988] using Mahler's method; unfortunately their proof had an unrecoverable gap.

Is it possible to say more about the diophantine properties of those numbers? How do they fit in Mahler's classification of transcendental numbers? How well can they be approximated by rationals?
Shallit conjectured [1999] that they are not Liouville numbers.

Liouville numbers

Let $\xi \in \mathbb{R}$. The irrationality measure of ξ is the supremum $\mu(\xi)$ of all $\tau \in \mathbb{R}$ such that there exist infinitely many $p/q \in \mathbb{Q}$ satisfying $\left|\xi - \frac{p}{q}\right| < \frac{1}{q^\tau}$.

If $\xi \in \mathbb{Q}$, then $\mu(\xi) = 1$; otherwise $\mu(\xi) \geq 2$ by the theory of continued fractions. If ξ is algebraic irrational, then $\mu(\xi) = 2$ by Roth's theorem. Actually $\mu(\xi) = 2$ for almost every $\xi \in \mathbb{R}$ [Khintchine 1924].

A Liouville number [Liouville 1844] is a real number with infinite irrationality measure. It can therefore be very well approximated by rationals.

Automatic numbers are not Liouville numbers

Theorem [Adamczewski-Cassaigne 2006]
Let ξ be a k-automatic number in base b. Then $\mu(\xi) \leq dk(k^m + 1)$, where $d = \#B$ is the cardinality of the internal alphabet of $\mathbf{x}$, and $m = \#N_k(\mathbf{x})$ is the cardinality of the k-kernel of $\mathbf{x}$.
In short, $\mu(\xi)$ can be effectively bounded knowing the automaton of $\mathbf{x}$. As a consequence, ξ is not a Liouville number.

The k-kernel of a sequence $\mathbf{u} = (u_n)_{n \geq 0}$ is defined as the set $N_k(\mathbf{u})$ of all sequences $(u_{k^i . n+j})_{n \geq 0}$, where $i \geq 0$ and $0 \leq j < k^i$. An infinite word is k-automatic iff its k-kernel is finite [Eilenberg 1974].

Combinatorial lemma

Lemma

Let $\mathbf{x}$ be a non-eventually periodic k-automatic infinite word on A. Let $u \in A^*$, $v \in A^+$, and $w \in A^*$ be such that uvw is a prefix of $\mathbf{x}$ and w is a prefix of vw. Let $m = \# N_k(\mathbf{x})$. Then $\frac{|uvw|}{|uv|} < k^m$.

Sketch of the proof of the lemma (1)

$(h, p, l) \in \mathbb{N}^3$ is admissible (with respect to $\mathbf{x}$) if:

(i) $1 \leq p \leq h \leq l$;

(ii) for $n = h, \ldots, l-1$, $x_{n-p} = x_n$;

(iii) $x_{l-p} \neq x_l$.

If u, v, w are as in the lemma, with w of maximal length, then $(|uv|, |v|, |uvw|)$ is admissible.

Assume that (h, p, l) is admissible and $l \geq hk^m$.

Sketch of the proof of the lemma (2)

For $0 \leq i \leq m$, we define $\mathbf{x}^{(i)} \in N_k(\mathbf{x})$ and (h_i, p, l_i) admissible with respect to $\mathbf{x}^{(i)}$.

We start with $\mathbf{x}^{(0)} = \mathbf{x}$, $h_0 = h$, and $l_0 = l$.

Then, given $\mathbf{x}^{(i)}$ and (h_i, p, l_i), for $0 \leq i < m$, let $l_{i+1} = \lfloor l_i/k \rfloor$, $r_i = l_i - kl_{i+1}$, $\mathbf{x}^{(i+1)} = (u^{(i)}_{kn+r_i})_{n \geq 0}$, and $h_{i+1} = l_{i+1} + p - \lfloor (l_i + p - h_i)/k \rfloor$. Note that (l_i) is strictly decreasing, with $l_i \geq h_0 k^{m-i}$, and (h_i) is nondecreasing.

Since $\# N_k(\mathbf{x}) = m$, we have $\mathbf{x}^{(i)} = \mathbf{x}^{(j)}$ for some $i < j$.
(h_i, p, l_i) and (h_j, p, l_j) are admissible for the same sequence, with $h_i \leq h_0 \leq h_0 k^{m-j} \leq l_j < l_i$, a contradiction.

Sketch of the proof of the theorem (1)

We construct a family of rational numbers p_n/q_n such that

$$\frac{1}{q_n^{k^m+1+\varepsilon}} < \left|\xi - \frac{p_n}{q_n}\right| < \frac{1}{q_n^{1+1/d}}$$

where the upper bound comes from the construction (pigeon-hole principle in the internal alphabet) and the lower bound from the lemma. Moreover, we ensure that $q_n < q_{n+1} < q_n^{k+\varepsilon}$.

Sketch of the proof of the theorem (2)

Assume that p/q is a rational number such that $\left|\xi - \frac{p}{q}\right| < \frac{1}{q^{M+\delta}}$, where $M = dk(k^m + 1)$ and $\delta > 0$.

We choose n such that $q_{n-1} \leq (2q)^d < q_n$.
Then by the triangle inequality

$$\left|\frac{p}{q} - \frac{p_n}{q_n}\right| \leq \left|\xi - \frac{p_n}{q_n}\right| + \left|\xi - \frac{p}{q}\right| < \frac{1}{q_n^{1+1/d}} + \frac{1}{q^{M+\delta}}$$

and we find that if q is large enough (and ε correctly adjusted) this is less than $\frac{1}{qq_n}$, implying that $p/q = p_n/q_n$, but the lower bound on $\left|\xi - \frac{p_n}{q_n}\right|$ contradicts the assumption on $\left|\xi - \frac{p}{q}\right|$.

Example: Thue-Morse numbers

Let $b \geq 2$ and $\mathbf{x}$ be the fixed point of the substitution $\theta : 0 \mapsto 01$, $1 \mapsto 10$, with the initial 0 deleted:

$$\mathbf{x} = 1101001100101101001011001101001\ldots$$

The associated real number θ_b is 2-automatic in base b, hence transcendental (this has been known for a long time [Mahler 1929]) and not a Liouville number.

Our theorem gives $\mu(\theta_b) \leq 20$. But a specific study yields $\mu(\theta_b) \leq 5$.

Generalizations

b-adic expansions can be replaced by β-expansions [Rényi 1957], where β is a Pisot or Salem number (real algebraic integer larger than 1, with all its conjugates of modulus at most 1). Most results that we have mentioned have an analogue in this setting. In particular, a number with an automatic β-expansion is either in $\mathbb{Q}(\beta)$, or it is transcendental but not a Liouville number.

Another possible generalization is to consider continued fraction expansions instead of b-adic expansions. If ξ has an automatic continued fraction expansion, we can conjecture that it is either rational (finite expansion), or quadratic (periodic expansion), or transcendental (aperiodic expansion), but this is not yet proved. Obviously such a number is not a Liouville number.