Jean-Paul ALLOUCHE

Thue, Combinatorics on words, and conjectures inspired by the Thue-Morse sequence

Thue, Combinatorics on words, and conjectures inspired by the Thue-Morse sequence

par JEAN-PAUL ALLOUCHE

On the occasion of Axel Thue's 150th birthday

RÉSUMÉ.   Nous décrivons quelques résultats récents sur la suite de Thue-Morse, ainsi que des questions ou conjectures, dont l'une, due à Shevelev, est résolue dans cet article.

ABSTRACT.   We describe some recent results on the Thue-Morse sequence. We also list open questions and conjectures, one of which is due to Shevelev and proved in this paper.

1. Introduction

The birth of combinatorics on words, i.e., the study of finite sequences (called *words*) taking their values in a finite set (called *alphabet*), aka the study of the free monoid generated by a finite set, can be dated to 1906, when the first of the two seminal papers [60, 61] of Axel Thue appeared. Thue was interested in constructing an infinite sequence on three symbols with no squares (i.e., without two consecutive identical blocks) in it, and/or an infinite sequence on two symbols with no cubes (i.e., without three consecutive identical blocks) in it. Thue's sequence on two symbols was actually already described in a number-theoretic paper by Prouhet [44] addressing what is now known as the Prouhet-Tarry-Escott problem (see, e.g., [15]), so that to some extent Prouhet might be considered as one of the fathers of combinatorics of words (see [20]).

To begin with we would like to point out the first few lines of Thue's second paper [61], where he writes (translation by Berstel in [14]):

For the development of logical sciences it will be important, without consideration for possible applications, to find large domains for speculation about difficult problems. In this paper, we present some investigations in the theory of sequences of symbols, a theory that has some connections with number theory.

Combinatorics on words developed extensively in the last forty years, interacting with many fields, from number theory to harmonic analysis, from theoretical computer science to physics, with incursions towards tiling, music, etc.

We will briefly indicate some of these developments as well as pointers to surveys or to the original bibliography. Then we will single out a few conjectures or open problems to show that the domain is still full of really nontrivial and interesting questions.

2. The Thue-Morse sequence and its numerous occurrences

2.1. Recalling the definition of the Thue-Morse sequence.

Definition. The Thue-Morse sequence is the sequence $\mathfrak{t} = (\mathfrak{t}(n))_{n \geq 0}$ defined by $\mathfrak{t}(n) = 0$ if the sum of the binary digits of the integer n is even, and $\mathfrak{t}(n) = 1$ if the sum of the binary digits of n is odd. We let $\mathfrak{u} = (\mathfrak{u}(n))_{n \geq 0}$ denote the sequence defined by $\mathfrak{u}(n) = (-1)^{\mathfrak{t}(n)}$.

Remark. A useful characteristic property of the sequence $\mathfrak{u}$ is the recurrence property

$$\mathfrak{u}(2n) = \mathfrak{u}(n) \ \text{ and } \ \mathfrak{u}(2n + 1) = -\mathfrak{u}(n).$$

Note that several equivalent definitions can be given for the sequences $\mathfrak{t}$ and $\mathfrak{u}$. In particular $\mathfrak{t}$ is the infinite fixed point, beginning with 0, of the (uniform) morphism $0 \to 01$, $1 \to 10$. For more about the sequences $\mathfrak{t}$ and $\mathfrak{u}$ we refer the reader to, e.g., [6].

After the papers of Thue [60, 61] on what is now called the Thue-Morse sequence (or the Prouhet-Thue-Morse sequence), the sequence was rediscovered (in particular by Morse [41]) and/or studied in very many different contexts. J. Shallit and the author wrote a survey some years ago entitled "The ubiquitous Thue-Morse sequence" [6], but since that time, we have gathered more and more new references on occurrences of this sequence in the literature. We would like to briefly allude here to only two "new" directions.

2.2. Nonrepetitive coloring of graphs. The seminal paper [9] introduced the notion of coloring graphs "without repetitions": a coloring of the set of edges of a graph G is called non-repetitive if the sequence of colors on any path in G is non-repetitive, i.e., has no squares (consecutive identical blocks) in it. In the last ten years a very large number of papers have been devoted to this subject, and it would deserve a full survey by itself.

2.3. The Thue-Morse sequence in games, economics, etc. Suppose that two players A and B are playing a sequential tournament where the first mover may have some advantage (or disadvantage) just by playing first. Then, if the players are to play several consecutive games, it seems fair that the order be, e.g., *A B A B A B* ... In order to try to improve the fairness of the tournament, one can imagine that the first player plays, then each player –begining with the second one– plays twice (this is for example the case for tiebreak in tennis), yielding the sequence *A B B A A B B A A B B A* ... Trying to improve again and again on this alternation leads to... the Thue-Morse sequence: to cite but a few papers on this subject, see [16] cited in [37]; also see [25, 36, 43, 45].

3. Some conjectures and open questions

3.1. A recent conjecture of Shevelev. V. Shevelev wrote recently several papers on the Thue-Morse sequence or related problems, see [47, 48, 49, 50, 51, 52, 53, 54, 55, 56], also see the two papers by Moses and Shevelev [57, 58]. These papers contain nice results, but also questions. One of Shevelev's questions about the Thue-Morse sequence is rather intriguing [56]. In order to give the flavor of this question, let us recall that an integer is called *evil* [in French *païen*] (resp. *odious* [in French *impie*]) if the sum of its binary digits is even (resp. odd) I learned in the paper of Fraenkel [33] that the terminology, inspired phonetically by the words "even" and "odd", was coined by the authors of [11] when they were composing their book[1], see [12, p. 463]. A particular case of Shevelev's question is: is it true that for any odd integer a, evil and odious numbers alternate in the increasing sequence of integers n such that $\mathfrak{u}(n+a) = \mathfrak{u}(n)$ where $\mathfrak{u}$ is the ± 1 Thue-Morse sequence?

More generally Shevelev made the following conjecture that we will prove here.

Theorem 3.1 (Shevelev's conjecture). *Let $(\mathfrak{u}(n))_{n\geq 0} = ((-1)^{\mathfrak{t}(n)})_{n\geq 0}$. Let a be a positive integer. Let $B_a := \{\ell_0 < \ell_1 < \ell_2 < \ldots\}$ (resp. $C_a = \{m_0 < m_1 < m_2 < \ldots\}$) denote the sequences of integers in increasing order satisfying $\mathfrak{u}(\ell + a) = -\mathfrak{u}(\ell)$, resp. $\mathfrak{u}(m + a) = \mathfrak{u}(m)$.*
Let $(\beta_a(n))_{n\geq 0}$ (resp. $(\gamma_a(n))_{n\geq 0}$) be the ± 1 sequences defined by

$$\beta_a(n) := \mathfrak{u}(\ell_n), \ \gamma_a(n) := \mathfrak{u}(m_n).$$

Then the sequences β_a and γ_a are periodic, of smallest period $2^{v(a)+1}$ where $v(a)$ is the 2-adic valuation of a (i.e., the largest integer k such that 2^k divides a). They satisfy $\beta_a = -\gamma_a$. Furthermore the prefix of length $2^{v(a)+1}$

---

[1]In the papers [32, 33] the terms "vile" and "dopey" are proposed for other kinds of integers. We also cite a suggestion of Ingrid Daubechies to call *perfidy* the fact of being evil or odious, as *parity* is the fact of being even or odd, see http://blog.tanyakhovanova.com/?p=97

of the sequence γ_a is equal to the prefix of length $2^{v(a)+1}$ of the Thue-Morse sequence $\mathfrak{u}$ if $\mathfrak{u}(a) = 1$ and to "minus this prefix" (i.e., where all $+1$ are replaced by -1 and conversely) if $\mathfrak{u}(a) = -1$.

Proof. We first prove the periodicity assertion for sequence γ_a. The proof splits into two parts depending on the parity of a.

– Suppose that a is odd. Note that, in order to prove the assertions "equal to the prefix..." and "equal to minus the prefix...", it suffices to prove the assertion "equal to ± 1 the prefix..." thanks to the remark that the sequence γ_a begins with 1 when $\mathfrak{u}(a) = 1$ and in -1 when $\mathfrak{u}(a) = -1$. Namely, we first note that for any a (odd or even) such that $\mathfrak{u}(a) = 1$, it is clear that $\min C_a = 0$, hence $\mathfrak{u}(\min C_a) = \mathfrak{u}(0) = 1$, thus $\gamma_a(0) = 1$. Now $(\mathfrak{u}(0), \mathfrak{u}(1), \mathfrak{u}(2)) = (+1, -1, -1)$; also, for a is odd, $a + 1$ is even, thus $\mathfrak{u}(a + 1)$ and $\mathfrak{u}(a + 2)$ have opposite signs: namely, if $a = 2b + 1$, then $\mathfrak{u}(a+1) = \mathfrak{u}(2b+2) = \mathfrak{u}(b+1)$, while $\mathfrak{u}(a+2) = \mathfrak{u}(2b+3) = \mathfrak{u}(2(b+1)+1) = -\mathfrak{u}(b+1)$. Hence $(\mathfrak{u}(a), \mathfrak{u}(a+1), \mathfrak{u}(a+2)) = (\mathfrak{u}(a), +1, -1)$ or $(\mathfrak{u}(a), -1, +1)$. If $\mathfrak{u}(a) = -1$, we have $\min C_a = 2$ or 1, thus $\mathfrak{u}(\min C_a) = -1$. In other words in the case where a is odd we always have that $\gamma_a(0) = \mathfrak{u}(\min C_a) = \mathfrak{u}(a)$. Hence, what we have to prove in the case a odd is that γ_a is equal to one of the sequences $(+1 \ -1)^\infty$ or $(-1 \ +1)^\infty$.

Now consider the sequence $(w_a(n))_{n \geq 0}$ where $w_a(n) := \mathfrak{u}(a + n) + \mathfrak{u}(n)$. Clearly $w_a(n) = \pm 2$ if n belongs to C_a, and $w_a(n) = 0$ otherwise. More precisely $w_a(n) = 2$ if n belongs to C_a and $\mathfrak{u}(n) = 1$, and $w_a(n) = -2$ if n belongs to C_a and $\mathfrak{u}(n) = -1$. In other words $(w_a(n))_{n \geq 0}$ is the sequence obtained by interspersing blocks of zeros into the sequence $(2\gamma_a(n))_{n \geq 0}$. What we have to prove is that the sequence $(2\gamma_a(n))_{n \geq 0}$ is one of the alternating sequences $(+2 \ -2)^\infty$ or $(-2 \ +2)^\infty$. Since $2\gamma_a$ takes only the values $+2$ and -2, this is equivalent to proving that the summatory function of sequence w_a either takes only the values $0, +2$, or only the values $0, -2$. Now the summatory function of w_a is the sequence of coefficients of the formal power series $S_a(X) := \frac{1}{1-X} \sum_{n \geq 0} w_a(n) X^n$. We have

$$
\begin{aligned}
S_a(X) &= \frac{1}{1-X} \sum_{n \geq 0} (\mathfrak{u}(n+a) + \mathfrak{u}(n)) X^n \\
&= \frac{1}{1-X} \left(\sum_{m \geq a} \mathfrak{u}(m) X^{m-a} + \sum_{n \geq 0} \mathfrak{u}(n) X^n \right).
\end{aligned}
$$

Hence, multiplying by X^a,

$$
X^a S_a(X) = -\frac{1}{1-X} \left(\sum_{0 \leq m \leq a-1} \mathfrak{u}(m) X^m \right) + \frac{1 + X^a}{1 - X} \sum_{n \geq 0} \mathfrak{u}(n) X^n.
$$

But

$$
\begin{aligned}
\sum_{n\geq 0} \mathfrak{u}(n)X^n &= \sum_{n\geq 0}\mathfrak{u}(2n)X^{2n} + \sum_{n\geq 0}\mathfrak{u}(2n+1)X^{2n+1}\\
&= \sum_{n\geq 0}\mathfrak{u}(n)X^{2n} - \sum_{n\geq 0}\mathfrak{u}(n)X^{2n+1}\\
&= (1-X)\sum_{n\geq 0}\mathfrak{u}(n)X^{2n}
\end{aligned}
$$

(this is well-known, as is the iteration yielding that $\sum_{n\geq 0}\mathfrak{u}(n)X^n$ equals $\prod_{n\geq 0}(1-X^{2^n})$). Thus

$$
(*) \quad X^a S_a(X) = \frac{-1}{1-X}\left(\sum_{0\leq m\leq a-1}\mathfrak{u}(m)X^m\right) + (1+X^a)\sum_{n\geq 0}\mathfrak{u}(n)X^{2n}.
$$

Now recall that a is odd, say $a = 2b+1$. Then,

$$
\begin{aligned}
\sum_{0\leq m\leq a-1}\mathfrak{u}(m)X^m &= \sum_{0\leq m\leq 2b}\mathfrak{u}(m)X^m\\
&= \sum_{0\leq m\leq b}\mathfrak{u}(2m)X^{2m} + \sum_{0\leq m\leq b-1}\mathfrak{u}(2m+1)X^{2m+1}\\
&= \sum_{0\leq m\leq b}\mathfrak{u}(m)X^{2m} - \sum_{0\leq m\leq b-1}\mathfrak{u}(m)X^{2m+1}\\
&= \mathfrak{u}(b)X^{2b} + (1-X)\left(\sum_{0\leq m\leq b-1}\mathfrak{u}(m)X^{2m}\right).
\end{aligned}
$$

Using $(*)$ we have

$$
X^a S_a(X) = \frac{-1}{1-X}\left(\sum_{0\leq m\leq 2b}\mathfrak{u}(m)X^m\right) + (1+X^{2b+1})\sum_{n\geq 0}\mathfrak{u}(n)X^{2n}
$$

hence

$$
X^a S_a(X) = -\frac{\mathfrak{u}(b)X^{2b}}{1-X} - \sum_{0\leq m\leq b-1}\mathfrak{u}(m)X^{2m} + (1+X^{2b+1})\sum_{n\geq 0}\mathfrak{u}(n)X^{2n}.
$$

Thus

$$
(**) \quad X^a S_a(X) = -\mathfrak{u}(b)X^{2b}\sum_{j\geq 0}X^j - \sum_{0\leq m\leq b-1}\mathfrak{u}(m)X^{2m}
$$
$$
+ (1+X^{2b+1})\sum_{n\geq 0}\mathfrak{u}(n)X^{2n}.
$$

In order to finish the proof of the case a odd, we have to prove that all the coefficients of $S_a(X)$ either take only the values 0, 2, or take only the values 0, -2. Of course it suffices to prove the same claim for the coefficients of $X^a S_a(X)$. Write $X^a S_a(X) = \sum c_k X^k$. Looking at $(**)$ we see that

- for $k \leq a-1$, we have clearly $c_k = 0$;

- for $2k \geq a$, we have $c_{2k} = -\mathfrak{u}(b) + \mathfrak{u}(k) \in \{-\mathfrak{u}(b) \pm 1\}$;
- for $2k + 1 \geq a$, we have $c_{2k+1} = -\mathfrak{u}(b) + \mathfrak{u}(k - b) \in \{-\mathfrak{u}(b) \pm 1\}$.

Hence, either $\mathfrak{u}(b) = +1$ and all coefficients of $X^a S_a(X)$ belong to $\{-2, 0\}$, or $\mathfrak{u}(b) = -1$ and all coefficients of $X^a S_a(X)$ belong to $\{0, 2\}$, and we are done.

– Suppose that a is even. We note that in order to address this case it suffices to prove that if the statement in Theorem 3.1 is true for some integer $a \geq 1$, then it is true for $2a$. Recall that the Thue-Morse sequence can be generated by iteratively applying to $+1$ the morphism σ defined on $\{+, -\} := \{+1, -1\}$ by $\sigma(+) = + -$, $\sigma(-) = - +$. Thus the prefix of length 2^{d+1} of the Thue-Morse sequence is equal to the image by σ of the prefix of length 2^d. Hence it suffices to prove that for any $a \geq 1$ one has $C_{2a} = 2C_a \cup (2C_a + 1)$: namely this implies that the $2n$-th term of C_{2a} is $2m_n$ where m_n is the n-therm of C_a, and that the $2n + 1$-st term of C_{2a} is $2m_n + 1$; we can then write $\gamma_{2a}(2n) = \mathfrak{u}(2m_n) = \mathfrak{u}(m_n) = \gamma_a(n)$ and $\gamma_{2a}(2n + 1) = \mathfrak{u}(2m_n + 1) = -\mathfrak{u}(m_n) = -\gamma_a(n)$. But n belongs to C_{2a} if and only $\mathfrak{u}(n + 2a) = \mathfrak{u}(n)$. This happens if and only if either $n = 2k$ and $\mathfrak{u}(2k + 2a) = \mathfrak{u}(2k)$, or $n = 2k + 1$ and $\mathfrak{u}(2k + 1 + 2a) = \mathfrak{u}(2k + 1)$. This is equivalent to either $n = 2k$ and $\mathfrak{u}(k + a) = \mathfrak{u}(k)$, or $n = 2k + 1$ and $-\mathfrak{u}(k + a) = -\mathfrak{u}(k)$. This is exactly saying that n belongs to C_{2a} if and only if either n belongs to $2C_a$ or n belongs to $2C_a + 1$ (note that these sets are, of course, disjoint).

– To finish the proof, we follow exactly the same steps, with $\mathfrak{u}(n+a)+\mathfrak{u}(n)$ replaced by $\mathfrak{u}(n + a) - \mathfrak{u}(n)$. Note that this gives in passing the fact that $\beta_a = -\gamma_a$.

Remark. Shevelev proved in [56] his conjecture for the case where $a = 2^r$. The case $a = 1$ was actually proven by Bernhardt [13] (also see [10]). It was also given by P. Deléham (see his comment dated March 16 2004 in [42, sequence A003159]). Note that this case is related to the properties of the period-doubling sequence $(\mathfrak{z}(n))_{n \geq 0}$. The period-doubling sequence is defined as the infinite fixed point of the morphism defined on $\{+, -\} = \{+1, -1\}$ by $- \to - +$, $+ \to - -$. This morphism occurs, in particular, in the study of iterations of unimodal continuous functions. It is not difficult to see that $\mathfrak{u}(n) = \prod_{0 \leq k \leq n-1} \mathfrak{z}(k)$ (where, as usual, an empty product is equal to $+1$). Thus C_1 is the set of n's such that $\mathfrak{z}(n) = 1$.

The first values of the sequences B_a and C_a for small values of a's are given in the On-Line Encyclopedia of Integer Sequences [42], e.g., $C_1 = $ A079523; $C_2 = $ A081706; $C_3 = $ A161579; $C_4 = $ A161627; $C_5 = $ A161817; $C_6 = $ A161824; $C_7 = $ A162311; $C_8 = $ A161639; $C_9 = $ A161890.

There is a straightforward generalization of Shevelev's question. Let $(z(n))_{n \geq 0}$ be a ± 1 sequence taking each of the values ± 1 infinitely often. For each integer $a \geq 1$, let $B_a := \{\ell_0 < \ell_1 < \ell_2 < \ldots\}$ (resp. $C_a = \{m_0 < m_1 < m_2 < \ldots\}$) denote the sequence of integers in increasing order satisfying $z(\ell + a) = -z(\ell)$ (resp. $z(m + a) = z(m)$). These two sequences of integers reflect how different sequence u and each of its shifted sequences are. Now evaluate this difference in terms of the sequence u iteself: let $(\beta_a(n))_{n \geq 0}$ (resp. $(\gamma_a(n))_{n \geq 0}$) be the ± 1 sequences defined by $\beta_a(n) := z(\ell_n)$ (resp. $\gamma_a(n) := z(m_n)$). What can be said about sequences β_a and γ_a? Or for which sequences u do the sequences β_a and γ_a have interesting properties?

3.2. The zeros of the Thue-Morse Dirichlet series.

In a 1985 paper [4] H. Cohen and the author looked at the Dirichlet series $\sum_{n \geq 0} \frac{u(n)}{(n+1)^s}$ a priori defined for $\Re s > 1$. They proved in particular that this series admits an analytic continuation to the whole plane. This continuation admits "trivial zeros" (all non-positive integers), and "non-trivial zeros" (the complex numbers $2ik\pi / \log 2$ for k integer). This suggests the following Riemann-like hypothesis.

Question. Is it true that $\sum_{n \geq 0} \frac{u(n)}{(n+1)^s}$ has no other non-trivial zeros than the complex numbers $2ik\pi / \log 2$ for k integer?

Remark. It might well be that this question is very difficult to answer. It might also well be the case that the answer would have no interest at all: a proof of the Riemann hypothesis for the zeta function would give precious information on the prime numbers, but we do not know of any consequence of the Riemann-like hypothesis above.

3.3. Looking for a "simple" expression for a certain infinite product.

The following infinite product and its simple expression were given in [62, 46]

$$P := \prod_{n \geq 0} \left(\frac{2n+1}{2n+2} \right)^{u(n)} = \frac{\sqrt{2}}{2}.$$

(Note that this product and similar products in this section are easily proven convergent by using Abel's "summation" for products.) While several generalizations can be found in the literature (see, e.g., [4, 5, 8, 38]), no "simple" value is known for a very similar product Q given below. Namely a strange product appears in [31, p. 193]:

$$R := \prod_{n \geq 1} \left(\frac{(4n+1)(4n+2)}{4n(4n+3)} \right)^{u(n)}.$$

Actually it is easily proven that $R = \frac{3}{2Q}$, where

$$Q := \prod_{n \geq 1} \left(\frac{2n}{2n+1} \right)^{\mathfrak{u}(n)}.$$

Question. Does the infinite product Q above have a "simple" value? Is it a transcendental number? Is the Flajolet-Martin constant φ defined by $\varphi := 2^{-1/2} e^{\gamma} \frac{2}{3} R = 0.77351...$ transcendental?

Remark. The author gave an easy proof for the value of P (which was written down in [7] and [8]) by computing the product PQ. We do not resist giving an even easier proof. Using the relations $\mathfrak{u}(2n) = \mathfrak{u}(n)$ and $\mathfrak{u}(2n+1) = -\mathfrak{u}(n)$, we can write

$$
\begin{aligned}
P &= \prod_{n \geq 0} \left(\frac{2n+1}{2n+2} \right)^{\mathfrak{u}(n)} & &= \left(\prod_{n \geq 0} \left(\frac{2n+1}{2n+2} \right)^{\mathfrak{u}(2n+1)} \right)^{-1} \\
&= \left(\frac{\prod_{n \geq 1} \left(\frac{n}{n+1} \right)^{\mathfrak{u}(n)}}{\prod_{n \geq 1} \left(\frac{2n}{2n+1} \right)^{\mathfrak{u}(2n)}} \right)^{-1} & &= \left(\frac{\prod_{n \geq 1} \left(\frac{n}{n+1} \right)^{\mathfrak{u}(n)}}{\prod_{n \geq 1} \left(\frac{2n}{2n+1} \right)^{\mathfrak{u}(n)}} \right) \\
&= \left(\prod_{n \geq 1} \left(\frac{2n+1}{2(n+1)} \right)^{\mathfrak{u}(n)} \right)^{-1}.
\end{aligned}
$$

Hence, multiplying by P,

$$P^2 = \left(\prod_{n \geq 1} \left(\frac{2n+1}{2(n+1)} \right)^{\mathfrak{u}(n)} \right)^{-1} \prod_{n \geq 0} \left(\frac{2n+1}{2n+2} \right)^{\mathfrak{u}(n)} = \frac{1}{2}.$$

Thus

$$\prod_{n \geq 0} \left(\frac{2n+1}{2n+2} \right)^{\mathfrak{u}(n)} = \frac{1}{\sqrt{2}}.$$

3.4. Looking for "another" proof of Cobham's theorem. Recall that a sequence $(v(n))_{n \geq 0}$ is called d-automatic for some integer $d \geq 2$ if the set of subsequences $\{(u(d^k n + j))_{n \geq 0}, \ k \geq 0, \ j \in [0, d^k - 1]\}$ is finite (see, e.g., [7]). For example the Thue-Morse sequence is 2-automatic (it is also 2^r-automatic for any $r \geq 1$). Cobham's theorem asserts that a sequence which is both q- and r-automatic, where q and r are multiplicatively independent (i.e., $\log q / \log r$ irrational), must be ultimately periodic [24]. The proof of Cobham is very technical, as are more recent proofs. The quest for a "simple" –or at least more "conceptual"– proof whose early start was [34] is still open.

3.5. Almost everywhere automatic sequences. Deshouillers, whose motivation was the results in [26, 27], asked the following question [28]. Call a sequence $(v(n))_{n\geq0}$ *almost everywhere d-automatic* if there exists a d-automatic sequence $(w(n))_{n\geq0}$ such that v and w are equal almost everywhere (i.e., such that the set of n for which $u(n) \neq v(n)$ has natural density 0, i.e., $\lim_{N\to\infty} \frac{1}{N}\sharp\{n \leq N,\ v_n \neq w_n\} = 0$). Define similarly *almost everywhere periodic* sequences, *almost everywhere ultimately periodic sequences*, and *almost everywhere constant* sequences. A simple example is the characteristic function of squares: this sequence is almost everywhere constant (actually almost everywhere equal to 0).

Question. Is the following generalization of Cobham's theorem true? Let d_1 and d_2 be two multiplicatively independent integers (i.e., $\log d_1 / \log d_2$ is irrational). If the sequence $(v(n))_{n\geq0}$ is both almost everywhere d_1-automatic and almost everywhere d_2-automatic, then it is almost everywhere ultimately periodic.

3.6. Sum of digits, pseudo-randomness, distribution modulo 1. The Thue-Morse sequence or its version not reduced modulo 2 (i.e., the sum of binary digits of the integers) was used in several questions related to pseudo-randomness (in the sense of Bass or in the sense of Bertrandias), see, e.g., the paper of Mendès France (see [40] and the references therein), or to distribution modulo 1. The subject continues to be explored, see, e.g., [1]. A nice question of Gelfond was recently answered in [39].

Theorem 3.2 (Mauduit-Rivat)**.** *The sum of q-ary digits of the prime numbers is uniformly distributed in the non-trivial arithmetic progressions.*

3.7. Algebraic independence of power series on a finite field. Let p be a prime number. Christol's theorem (see [21, 23]) asserts that the formal power series $\sum a_n X^n$ in $\mathbb{Z}/p\mathbb{Z}[[X]]$ is algebraic over $\mathbb{Z}/p\mathbb{Z}(X)$ if and only if the p-kernel of the sequence $(a_n)_{n\geq0}$, i.e., the set of subsequences $\{(a_{p^k n+r})_{n\geq0},\ k \geq 0,\ r \in [0, p^k - 1]\}$ is finite. In other words a combinatorial property of the sequence $(a_n)_{n\geq0}$ is equivalent to the algebraicity of the associated formal power series. A tempting question is then whether there exists some combinatorial property of two sequences $(a_n)_{n\geq0}$ and $(b_n)_{n\geq0}$ that is equivalent to the property that the formal power series $\sum a_n X^n$ and $\sum b_n X^n$ are algebraically dependent over $\mathbb{Z}/p\mathbb{Z}(X)$. Such a condition could be the finiteness of some "kernel" of sequence $(a_n)_{n\geq0}$, where the extracted subsequences would somehow depend on sequence $(b_n)_{n\geq0}$. G. Christol told us (private communication) that he does not believe that such a condition can exist in general. If so, could we find subcases where it would be possible to find such a combinatorial condition? A less ambitious question

would be to look at the Thue-Morse sequence $\mathbf{t}$ and to prove, e.g., that the series $\sum \mathbf{t}(W_1(n))X^n, \sum \mathbf{t}(W_2(n))X^n, \ldots, \sum \mathbf{t}(W_r(n))X^n$ modulo 2 are algebraically independent over $\mathbb{Z}/2\mathbb{Z}(X)$, where the W_k's are polynomials of degree k with rational coefficients such that $W(\mathbb{N}) \subset \mathbb{N}$. (For the linear independence of such formal power series over $\mathbb{Z}/2\mathbb{Z}(X)$, see [17].)

3.8. D-finite formal power series and automatic sequences. The formal power series $\sum_{n\geq 0} \mathbf{t}(n)X^n$ considered as an element of $\mathbb{Q}[[X]]$ is transcendental over $\mathbb{Q}(X)$. This is, e.g., a consequence of a theorem of Fatou [30]: *a power series $\sum_{n\geq 0} a_n X^n$ with integer coefficients that converges inside the unit disk is either rational or transcendental over $\mathbb{Q}(X)$.* Actually more is known: the formal power series $\sum_{n\geq 0} \mathbf{t}(n)X^n$ being irrational and having its radius of convergence equal to 1, it admits the unit circle as natural boundary (from a theorem of Carlson [19] extending Fatou's). Another proof of this transcendence using Christol's theorem can be given. If the formal power were algebraic, then the series $\sum_{n\geq 0}(\mathbf{t}_n \bmod 2)X^n$ (resp. the series $\sum_{n\geq 0}(\mathbf{t}_n \bmod 3)X^n$) would be algebraic over $\mathbb{F}_2(X)$ (resp. $\mathbb{F}_3(X)$). Hence, the sequence $(\mathbf{t}_n)_{n\geq 0}$ with values $0, 1$ would be both 2-automatic and 3-automatic. From Cobham's theorem [24] the sequence $(\mathbf{t}_n)_{n\geq 0}$ would thus be ultimately periodic which is not true (recall that $(\mathbf{t}_n)_{n\geq 0}$ does not contain cubes).

More generally, if a formal power series $\sum a_n X^n$ has all its coefficients in the field of rational numbers $\mathbb{Q}$, and is algebraic of degree d over $\mathbb{Q}(X)$, we know, as a consequence of Eisenstein's theorem (announced by Eisenstein in [29], and proved by Heine in [35]) that the set of prime numbers $\mathcal{P}$ that divide the denominator of at least one coefficient a_n is finite. Hence it makes sense to reduce the series modulo any prime not in $\mathcal{P}$. The reduction modulo such a prime p is clearly algebraic over the field $\mathbb{F}_p(X)$ of degree $d_p \leq d$. Hence from Christol's theorem, the sequence $(a_n \bmod p)_{n\geq 0}$ is p-automatic. A strategy for proving that the series $\sum a_n X^n$ is transcendental over $\mathbb{Q}(X)$ can thus be to prove that either it is algebraic of degree d_p for all but finitely many primes p, but that the d_p's are unbounded, or to prove that for some prime p that does not divide any of the denominators of the a_n's the sequence $(a_n \bmod p)_{n\geq 0}$ is not p-automatic. Several examples of this strategy are given, e.g., in [2].

Now, a notion generalizing algebraicity for power series is the notion of *D-finiteness*, also called *holonomy*. A very good reference is [59]. A formal power series is called *differentiably finite* (D-finite for short) or *holonomic* if it satisfies a linear differential equation with polynomial coefficients. Examples are the exponential series $f(X) = e^X$ that satisfies $f' = f$ or the series $g(X) = \log(1 + X)$ that satisfies $(1 + X)g'(X) = 1$. It is not hard to see

that any algebraic formal power series is D-finite. This notion makes sense in zero characteristic (in characteristic p the pth derivative of any formal power series is equal to 0). A natural question is thus whether there is a way to prove that a formal power series with say integer coefficients is not D-finite by proving that its projections modulo prime numbers do or do not satisfy some specific properties? The following question is a conjecture due to Christol [22].

Question. It is true that a *globally bounded* D-finite formal power series is *globally automatic*? In other words, is it true that a formal power series $\sum a_n X^n$ belonging to $\mathbb{Q}[[X]]$ having a non-zero radius of convergence as a series on $\mathbb{C}$ and for which there exist α, β in $\mathbb{Q}$ such that $\alpha \sum a_n(\beta X)^n$ belongs to $\mathbb{Z}[[X]]$ has the property that for all but finitely many primes p and for all positive integers h the sequence $(a_n \bmod p^h)_{n \geq 0}$ is p-automatic (or equivalently p^h-automatic)?

Remark. As noted by Christol the answer to the above question is yes if two classical conjectures respectively due to Bombieri and Dwork are true (see precise formulations in [22]).

3.9. Transcendence of morphic real numbers and morphic continued fractions. Several questions can be asked about the "transcendence" of the Thue-Morse sequence $\mathfrak{t}$. For example:

- is the real number $\sum_{n \geq 0} \mathfrak{t}(n)/2^n$ transcendental?
- is the formal power series $\sum_{n \geq 0} \mathfrak{t}(n)X^n \in \mathbb{Q}[[X]]$ transcendental over $\mathbb{Q}(X)$?
- is the formal power series $\sum_{n \geq 0}(\mathfrak{t}(n) \bmod 2)X^n \in \mathbb{Z}/2\mathbb{Z}[[X]]$ transcendental over $\mathbb{Z}/2\mathbb{Z}(X)$?
- is the formal power series $\sum_{n \geq 0}(\mathfrak{t}(n) \bmod 3)X^n \in \mathbb{Z}/3\mathbb{Z}[[X]]$ transcendental over $\mathbb{Z}/3\mathbb{Z}(X)$?
- is the continued fraction $[1+\mathfrak{t}(0), 1+\mathfrak{t}(1), 1+\mathfrak{t}(2), \ldots]$ transcendental over $\mathbb{Q}$?

Similar questions can be asked by replacing the Thue-Morse sequence (which, as recalled above, is the iterative fixed point beginning with 0 of the –uniform– morphism $0 \to 01$, $1 \to 10$) by iterative fixed points of –not necessarily uniform– morphisms, or even by morphic sequences, i.e., pointwise images of iterative fixed points of morphisms. A survey about these questions, and in particular the fact that the answers to the questions above about transcendence are respectively "yes, yes, no, yes, yes" can be found in [3]. More results that were either not published, or not proved at the time of that survey, are due in particular to Adamczewski and Bugeaud, and can be found in the nice book [18]. But, though we now know that "automatic real numbers" (real numbers whose expansion in some integer base is automatic, i.e., is the pointwise image of the iterative fixed point of

a uniform morphism) are either rational or transcendental, and that "automatic continued fractions" (i.e., continued fractions whose sequences of partial quotients are automatic) are either quadratic or transcendental, the general question about transcendence of real numbers whose expansion in some integer base, or whose continued fraction expansion, is the pointwise image of the iterative fixed point of some general morphism is still open.

4. Acknowledgments

We would like to thank V. Shevelev heartily for several interesting email discussions about the Thue-Morse sequence and about his conjecture. We also thank the referee for several detailed and useful comments.

References

[1] I. ABOU, P. LIARDET, *Mixing of Prouhet-Thue-Morse and Rudin-Shapiro sequences*, Annales Univ. Sci. Budapest., Sect. Comp. **40** (2013), 55–67.

[2] J.-P. ALLOUCHE, *Transcendence of formal power series with rational coefficients*, Theoret. Comput. Sci. **218** (1999), 143–160.

[3] J.-P. ALLOUCHE, *Automates et algébricités*, J. Théor. Nombres Bordeaux **17** (2005), 1–11.

[4] J.-P. ALLOUCHE, H. COHEN, *Dirichlet series and curious infinite products*, Bull. Lond. Math. Soc. **17** (1985), 531–538.

[5] J.-P. ALLOUCHE, J. SHALLIT, *Infinite products associated with counting blocks in binary strings*, J. Lond. Math. Soc. **39** (1989), 193–204.

[6] J.-P. ALLOUCHE, J. SHALLIT, *The ubiquitous Prouhet-Thue-Morse sequence*, in *Sequences and their Applications (Singapore, 1998)*, 1–16, Springer Ser. Discrete Math. Theor. Comput. Sci., Springer, London, 1999.

[7] J.-P. ALLOUCHE, J. SHALLIT, *Automatic Sequences. Theory, Applications, Generalizations*, Cambridge University Press, Cambridge, 2003.

[8] J.-P. ALLOUCHE, J. SONDOW, *Infinite products with strongly B-multiplicative exponents*, Annales Univ. Sci. Budapest., Sect. Comp. **28** (2008), 35–53.

[9] N. ALON, J. GRYTZUK, M. HAŁUSZCZAK, O. RIORDAN, *Non-repetitive colorings of graphs*, Random. Struct. Algor. **21** (2002), 336–346.

[10] R. A. BATES, H. MARURI-AGUILAR, E. RICCOMAGNO, R. SCHWABE, H. P. WYNN, *Self-avoiding generating sequences for Fourier lattice designs*, in *Algebraic Methods in Statistics and Probability II: Ams Special Session Algebraic Methods in Statistics and Probability, March 27-29, 2009, University Of Illinois at Urbana-Champaign*, Eds. M. A. G. Viana, H. P. Wynn, Contemporary Mathematics **516** (2010), 37–47.

[11] E. BERLEKAMP, J. CONWAY, R. GUY, *Winning Ways for Your Mathematical Plays*, vol. 1 and 2, Academic Press, 1982.

[12] E. BERLEKAMP, J. CONWAY, R. GUY, *Winning Ways for Your Mathematical Plays*, vol. 3, Second Edition, A. K. Peters, 2003.

[13] C. BERNHARDT, *Evil twins alternate with odious twins*, Math. Mag. **82** (2009), 57–62.

[14] J. BERSTEL, *Axel Thue's papers on repetitions in words: a translation*, Publications du LaCIM, Département de mathématiques et d'informatique **20**, Université du Québec à Montréal, 1995, 85 pages. Electronic version available at `http://lacim.uqam.ca/publications_pdf/20.pdf`

[15] P. BORWEIN, C. INGALLS, *The Prouhet-Tarry-Escott problem revisited*, Enseign. Math. **40** (1994), 3–27.

[16] S. J. BRAMS, A. D. TAYLOR, *The Win-Win Solution: Guaranteeing Fair Shares to Everybody*, Norton, New York, 1999.

[17] X. LE BRETON, *Linear independence of automatic formal power series*, Discr. Math. **306** (2006), 1776–1780.

[18] Y. Bugeaud, *Distribution Modulo One and Diophantine Approximation*, Cambridge Tracts in Mathematics **193**, Cambridge University Press, Cambridge, 2012.

[19] F. Carlson, *Über Potenzreihen mit ganzzahligen Koeffizienten*, Math. Zeitschift **9** (1921), 1–13.

[20] A. Černý, *On Prouhet's solution to the equal powers problem*, Theoret. Comput. Sci. **491** (2013), 33–46.

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

[22] G. Christol, *Globally bounded solutions of differential equation*, in Analytic number theory (Tokyo, 1988), Lecture Notes in Math., **1434**, Springer, Berlin, (1990), 45–64.

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

[24] A. Cobham, *On the base-dependence of sets of numbers recognizable by finite automata*, Math. Systems Theory **3** 1969, 186–192.

[25] J. Cooper, A. Dutle, *Greedy Galois games*, Amer. Math. Monthly **120** (2013), 441–451.

[26] J.-M. Deshouillers, I. Z. Ruzsa, *The least non zero digit of n! in base* 12, Publ. Math. Debrecen **79** (2011), 395–400.

[27] J.-M. Deshouillers, *A footnote to The least non zero digit of n! in base* 12, Unif. Distrib. Theory **7** (2012), 71–73.

[28] J.-M. Deshouillers, Private communication.

[29] G. Eisenstein, *Über eine allgemeine Eigenschaft der Reihenentwicklungen aller algebraischen Funktionen*, Berlin. Sitzber. (1852), 441–443.

[30] P. Fatou, *Séries trigonométriques et séries de Taylor*, Acta Math. **30** (1906), 335–400.

[31] P. Flajolet, G. N. Martin, *Probabilistic counting algorithms for data base applications*, J. Comput. Sys. Sci. **31** (1985), 182–209.

[32] A. S. Fraenkel, *Aperiodic subtraction games*, Electron. J. Combin. **18** (2011), Paper 19.

[33] A. S. Fraenkel, *The vile, dopey, evil and odious game players*, Discr. Math. **312** (2012), 42–46.

[34] G. Hansel, *À propos d'un théorème de Cobham*, in *Actes de la fête des mots*, D. Perrin Éd., GRECO de programmation, Rouen (1982).

[35] E. Heine, *Der Eisensteinsche Satz über Reihen-Entwicklung algebraischer Functionen*, J. Reine Angew. Math. **45** (1853), 285–302.

[36] C. Kuzmics, T. Palfrey, B. W. Rogers, *Symmetric play in repeated allocation games*, Preprint, Bielfeld, Institute of Mathematical Economics, **468** (2012), available at the URL `http://ssrn.com/abstract=2106108`

[37] L. Levine, K. E. Stange, *How to make the most of a shared meal: plan the last bite first*, Amer. Math. Monthly **119** (2012), 550–565.

[38] G. Louchard, H. Prodinger, *The largest missing value in a composition of an integer and some Allouche-Shallit-type identities*, J. Integer Seq. **16**, Article 13.2.2 (2013), 16

[39] C. Mauduit, J. Rivat, *Sur un problème de Gelfond : la somme des chiffres des nombres premiers*, Ann. of Math. (2) **171** (2010), 1591–1646.

[40] M. Mendès France, *Nombres normaux. Applications aux fonctions pseudo-aléatoires*, J. Analyse Math. **20** (1967), 1–56.

[41] M. Morse, *Recurrent geodesics on a surface of negative curvature*, Trans. Amer. Math. Soc. **22** (1921), 84–100.

[42] *On-Line Encyclopedia of Integer Sequences*, available electronically at `http://oeis.org`

[43] I. Palacios-Huerta, *Tournaments, fairness and the Prouhet-Thue-Morse sequence*, Economic inquiry **50** (2012), 848–849.

[44] E. Prouhet, *Mémoire sur quelques relations entre les puissances des nombres*, C. R. Acad. Sci. Paris **33** (1851), 225. See also `http://upload.wikimedia.org/wikipedia/commons/c/c2/NoteProuhet.png`

[45] R. M. Richman, *Recursive binary sequences of differences*, Complex Systems **13** (2001), 381–392.

[46] D. Robbins, *Solution to Problem E 2692*, Amer. Math. Monthly **86** (1979), 394-395.

[47] V. Shevelev, *On monotonic strengthening of Newman-like phenomenon on* $(2m + 1)$-*multiples in base* $2m$, Preprint (2007), available electronically at `http://arxiv.org/abs/0710.3177`

[48] V. Shevelev, *On the Newman sum over multiples of a prime with a primitive or semiprimitive root* 2, Preprint (2007), available electronically at `http://arxiv.org/abs/0710.1354`

[49] V. Shevelev, *Two digit theorems*, Preprint (2007), available electronically at `http://arxiv.org/abs/0710.0144`

[50] V. Shevelev, *New digit results and several problems*, Preprint (2007), available electronically at `http://arxiv.org/abs/0709.3821`

[51] V. Shevelev, *Two algorithms for evaluation of the Newman digit sum, and a new proof of Coquet's theorem*, Preprint (2007), available electronically at `http://arxiv.org/abs/0709.0885`

[52] V. Shevelev, *A conjecture on primes and a step towards justification*, Preprint (2007), available electronically at `http://arxiv.org/abs/0706.0786`

[53] V. Shevelev, *On excess of the odious primes*, Preprint (2007), available electronically at `http://arxiv.org/abs/0707.1761`

[54] V. Shevelev, *Generalized Newman phenomena and digit conjectures on primes*, Int. J. Math. Math. Sci. (2008) Art. ID 908045, 12.

[55] V. Shevelev, *Exact exponent in the remainder term of Gelfond's digit theorem in the binary case*, Acta Arith. **136** (2009), 91–100.

[56] V. Shevelev, *Equations of the form* $t(x+a) = t(x)$ *and* $t(x+a) = 1-t(x)$ *for Thue-Morse sequence*, Preprint 2009 and 2012, available electronically at `http://arxiv.org/abs/0907.0880`

[57] V. Shevelev, P. J. C. Moses, *A family of digit functions with large periods*, Preprint (2012), available electronically at `http://arxiv.org/abs/1209.5705`

[58] V. Shevelev, P. J. C. Moses, *Tangent power sums and their applications*, Preprint (2012), available electronically at `http://arxiv.org/abs/1207.0404`

[59] R. P. Stanley, *Differentiably finite power series*, European J. Combin. **1** (1980), 175–188.

[60] A. Thue, *Über unendliche Zeichenreihen*, Norske vid. Selsk. Skr. Mat. Nat. Kl. **7** (1906), 1–22. Reprinted in *Selected mathematical papers of Axel Thue*, T. Nagell, ed., Universitetsforlaget, Oslo, (1977), 139–158.

[61] A. Thue, *Über die gegenseitige Lage gleicher Teile gewisser Zeichenreihen, Norske vid. Selsk. Skr. Mat. Nat. Kl.* **1** (1912), 1–67. Reprinted in *Selected mathematical papers of Axel Thue*, T. Nagell, ed., Universitetsforlaget, Oslo, (1977), 413–478.

[62] D. R. Woods, *Elementary problem proposal E 2692*, Amer. Math. Monthly **85** (1978), 48.

Jean-Paul Allouche
Institut de Mathématiques de Jussieu-PRG
Équipe Combinatoire et Optimisation
Université Pierre et Marie Curie, Case 247
4 Place Jussieu
F-75252 Paris Cedex 05 France
E-mail: `jean-paul.allouche@imj-prg.fr`
URL: `http://www.imj-prg.fr/~jean-paul.allouche`