The ubiquitous Prouhet-Thue-Morse sequence

Jean-Paul Allouche[1] and Jeffrey Shallit[2]

[1] CNRS, LRI, Université Paris-Sud, Bâtiment 490
F-91405 Orsay Cedex, France
Email: allouche@lri.fr

[2] Department of Computer Science, University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
Email: shallit@uwaterloo.ca

Abstract. We discuss a well-known binary sequence called the Thue-Morse sequence, or the Prouhet-Thue-Morse sequence. This sequence was introduced by Thue in 1906 and rediscovered by Morse in 1921. However, it was already implicit in an 1851 paper of Prouhet. The Prouhet-Thue-Morse sequence appears to be somewhat ubiquitous, and we describe many of its apparently unrelated occurrences.

1 Introduction

In his seminal 1906 and 1912 papers [65, 66], the Norwegian mathematician Axel Thue (1863–1922) noted that any binary sequence of length ≥ 4 *must* contain a square, i.e., two consecutive identical blocks (the easy proof is left to the reader). He then asked whether it is possible to find an infinite sequence on three letters without squares, i.e., without two consecutive identical blocks. He also asked whether it is possible to find an infinite binary sequence that contains no cube, i.e., no three consecutive identical blocks, or even no overlap, i.e., no sub-block of the form $awawa$, where $a \in \{0, 1\}$ and w is a binary block. The answer to all three questions is positive. Thue used a sequence $\mathbf{t}$ whose construction is given in the next section,

$$\mathbf{t} = 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0 \cdots$$

It happens that this sequence $\mathbf{t}$ is really ubiquitous in the literature. In this paper we survey a few of its occurrences in combinatorics on words, differential geometry, number theory, iteration of continuous functions, and mathematical physics. Note that we do not give *all* properties of the sequence $\mathbf{t}$, but rather show how it occurred as a "natural" answer to various apparently unrelated questions.

2 Definition

We first give a formal definition of the Prouhet-Thue-Morse sequence.

Definition 1. *We denote by* $\mathbf{t} = (t_n)_{n\geq 0}$ *the* Prouhet-Thue-Morse *sequence over* $\{0,1\}$*, defined recursively by* $t_0 = 0$ *and* $t_{2n} = t_n$*,* $t_{2n+1} = \overline{t_n}$ *for all* $n \geq 0$*, where, for* $x \in \{0,1\}$*, we define* $\overline{x} = 1 - x$*.*

Denote by $s_k(n)$ the sum of the digits in the base-k representation of the integer n. Since we clearly have $s_2(2n) = s_2(n)$ and $s_2(2n+1) = s_2(n) + 1$ for every $n \geq 0$, we easily obtain the following equivalent definition:

Proposition 1. *The Prouhet-Thue-Morse sequence* $\mathbf{t}$ *is equal to the sequence* $(s_2(n) \bmod 2)_{n\geq 0}$*.*

Yet another definition, easily seen to be equivalent to the previous two, is the following:

Proposition 2. *Let* X *be an indeterminate. Then we have*

$$\begin{aligned}\prod_{i\geq 0}(1 - X^{2^i}) &= (1-X)(1-X^2)(1-X^4)\cdots \\ &= 1 - X - X^2 + X^3 + \cdots \\ &= \sum_{j\geq 0}(-1)^{t_j}X^j.\end{aligned}$$

3 Combinatorics on words

3.1 The pioneering work of Thue

Our first theorem is the one we mentioned in the introduction. It is due to Thue [65, 66].

Theorem 1 (Thue). *The Prouhet-Thue-Morse sequence* $\mathbf{t}$ *is overlap-free.*

For $n \geq 1$ let v_n be the number of 1's between the n'th and $(n+1)$'st occurrence of 0 in the sequence $\mathbf{t}$. Let $\mathbf{v} = (v_n)_{n\geq 1}$. Hence

$$\mathbf{v} = 2\ 1\ 0\ 2\ 0\ 1\ 2\ 1\ 0\ 1\ 2 \cdots$$

Thue proved, as a corollary of Theorem 1 above, the following:

Corollary 1. *The sequence* $\mathbf{v} = (v_n)_{n\geq 1}$ *is square-free.*

This work of Thue (see also Berstel [20] and references therein) was the starting point of an important branch of combinatorics, now called *combinatorics on words*. It is worth noting that Thue explained he had no particular application in mind, but he thought the problem was interesting enough in itself to deserve attention.

Thue's papers appeared in an obscure Norwegian journal, and for a long time were not widely known or appreciated. His original results were rediscovered by several different authors, including Marston Morse; see [38, 19].

Although there are uncountably many overlap-free sequences on two symbols, the Prouhet-Thue-Morse sequence is, roughly speaking, the "canonical" example. For example, if in addition to being overlap-free we add some extra requirement, then we often find that the only such sequence is the Prouhet-Thue-Morse sequence or a simple variant. For example, consider the following theorem of Berstel [20].

Theorem 2. *The lexicographically largest overlap-free binary sequence beginning with* 0 *is* $\mathbf{t}$*, the Prouhet-Thue-Morse sequence.*

Recently, together with J. Currie [10], we generalized this theorem, proving in particular the following result:

Theorem 3. *The lexicographically largest overlap-free binary sequence is the sequence* $110110\mathbf{t}$.

3.2 The problem of infinite play in chess

In a little-known 1929 paper, the Dutch chess grandmaster Machgielis (Max) Euwe (1901–1981, world champion 1935–1937) independently discovered the Prouhet-Thue-Morse sequence $\mathbf{t}$ and applied it to an interesting problem in chess [35].

The so-called German rule states that a draw occurs if the same sequence of moves occurs three times in succession. Euwe proved, using the cube-free property of $\mathbf{t}$, that under such a rule infinite games of chess are possible.

For example, one can take the Prouhet-Thue-Morse sequence $\mathbf{t} = t_0 t_1 t_2 \cdots$ and map each 0 to the sequence of four moves (Ng1-f3, Ng8-f6, Nf3-g1, Nf6-g8) and each 1 to the sequence of four moves (Nb1-c3, Nb8-c6, Nc3-b1, Nc6-b8). The resulting sequence of moves represents a legal infinite game of chess, and no draw occurs under the German rule.

Later, Morse rediscovered the same technique [49, 51].

3.3 Morphisms of the free monoid

Let A be an *alphabet*, i.e., a finite set of symbols. The set of *words* over A (i.e., blocks — or strings — of symbols of A) equipped with the operation of *concatenation* is denoted by A^*: this is the *free monoid* generated by A.

Definition 2. *Let A be an alphabet. Let A^* be the free monoid generated by A. A map $\sigma : A^* \to A^*$ is called a* morphism *if for all words x and y in A^* we have $\sigma(xy) = \sigma(x)\sigma(y)$. Such a map is called a* uniform morphism *if all the images by σ of elements of the set A have the same number of letters.*

Note that a morphism is defined by its values on the elements of A. Note also that the morphism σ can be extended to infinite sequences by *continuity* (the set of finite and infinite sequences being equipped with the topology of simple convergence). This means that, for an infinite sequence $\mathbf{z} = (z_n)_{n \geq 0}$, the sequence $\sigma(\mathbf{z})$ is defined by $\sigma(\mathbf{z}) = \sigma(z_0)\sigma(z_1)\sigma(z_2)\cdots$

Proposition 3. *Define the morphism* μ *on the alphabet* $\{0, 1\}$ *by* $\mu(0) = 01$, $\mu(1) = 10$. *Then the Prouhet-Thue-Morse sequence* $\mathbf{t}$ *is the unique fixed point of* μ *that begins with* 0.

Proof. We first note that if an infinite sequence is a fixed point of μ, and begins with 0, it must begin with $\mu(0)$. Since $\mu(0) = 01$, the sequence must begin with $\mu(01) = \mu(0)\mu(1)$, hence with $\mu^2(0)$. Iterating, this means that the sequence must begin with $\mu^k(0)$ for every $k \geq 0$. This proves uniqueness. Since $\mu(0)$ begins with 0, we have that $\mu^{k+1}(0)$ begins with $\mu^k(0)$ for every k. Hence the sequence of words $(\mu^k(0))_{k\geq 0}$ converges towards an infinite sequence, say $\mathbf{z} = (z_n)_{n\geq 0}$, that clearly is a fixed point of μ.

Now, for $x \in \{0, 1\}$, we have $\mu(x) = x\overline{x}$, where, as previously, $\overline{x} = 1 - x$. Since $\mathbf{z}$ is a fixed point of μ, we thus have for every $n \geq 0$, that $z_{2n} = z_n$ and $z_{2n+1} = \overline{z_n}$. Hence the sequence $\mathbf{z}$ is equal to the Prouhet-Thue-Morse sequence. □

Is it possible to build another binary sequence that is both overlap-free and generated by a morphism? The next theorem, due to Séébold [62], answers this question negatively. Another proof of this result was given by Berstel and Séébold in [21].

Theorem 4 (Séébold). *If an overlap-free binary sequence is a fixed point of a non-trivial morphism, then it is either equal to* $\mathbf{t}$, *the Prouhet-Thue-Morse sequence, or its complement* $\overline{\mathbf{t}} = (\overline{t_n})_{n\geq 0} = 1001011001101001\cdots$.

The Thue-Morse sequence is the prototype of a class of sequences called 2-*automatic sequences.* Roughly speaking, a sequence is k-automatic if its n'th term is generated by a finite-state machine which takes as input the base-k expansion of n. For more about this class of sequences, see, for example, [29, 33, 4]. For the general subject of combinatorics on words, see [43].

4 Differential geometry

The Prouhet-Thue-Morse sequence has the nice property that it exhibits regularity without being ultimately periodic. Morse rediscovered the sequence $\mathbf{t}$ in 1921 in connection with differential geometry [48]. He proved the following:

Theorem 5 (Morse). *On a surface of negative curvature, having at least two different normal segments, there exists a set of geodesics that are recurrent without being periodic, and this set has the power of the continuum.*

To prove this result, one of the steps was the following proposition, given as a lemma in [48, p. 95]. We say a sequence $\mathbf{a} = a_0a_1a_2\cdots$ is *uniformly recurrent* if for each finite block of symbols w occurring in $\mathbf{a}$ there exists an integer n such that for all i, the sub-block $a_{i+1}\cdots a_{i+n}$ contains an occurrence of w. We say a sequence $\mathbf{a} = a_0a_1a_2\cdots$ is *ultimately periodic* if there exist integers $p \geq 1$, $N \geq 0$ such that $a_i = a_{i+p}$ for all $i \geq N$.

Proposition 4. *There exists an infinite sequence over* $\{1, 2\}$ *which is uniformly recurrent but not ultimately periodic.*

The sequence that Morse gives is exactly $\mathbf{t}$, where 0's have been replaced by 1's and 1's by 2's.

5 Number theory

5.1 The Prouhet-Tarry-Escott problem

As already noted by Adler and Li [2], the sequence $\mathbf{t}$ appears implicitly in an 1851 paper of Prouhet [55]. Prouhet was interested in a problem that was also studied more than fifty years later by Tarry and Escott, and which is now known as the "Prouhet-Tarry-Escott" or "multigrades" problem.

Prouhet addressed the following question: is it possible to find a partition of the set $\{0, 1, 2, \ldots, 2^N - 1\}$ into two disjoint sets I and J, such that $\sum_{i \in I} i^k = \sum_{j \in J} j^k$ for $k = 0, 1, 2, \ldots, t$? Of course we take $0^0 = 1$, so that in particular the case $k = 0$ shows that I and J must have the same number of elements. Prouhet proved that such a partition is possible if $N = t + 1$.

Theorem 6 (Prouhet). *The Prouhet-Thue-Morse sequence* $\mathbf{t} = (t_n)_{n \geq 0}$ *has the following property. Define*

$$I = \{i \in \{0, 1, 2, 3, \ldots, 2^N - 1\} : \ t_i = 0\},$$
$$J = \{j \in \{0, 1, 2, 3, \ldots, 2^N - 1\} : \ t_j = 1\}.$$

Then for $0 \leq k \leq N - 1$ *we have*

$$\sum_{i \in I} i^k = \sum_{j \in J} j^k.$$

For example, we have

$$0^k + 3^k + 5^k + 6^k + 9^k + 10^k + 12^k + 15^k = 1^k + 2^k + 4^k + 7^k + 8^k + 11^k + 13^k + 14^k$$

for $k = 0, 1, 2, 3$.

Prouhet actually studied the more general problem of finding a partition of $[0, q^N - 1]$ into q sets $I_1, I_2, \ldots, I_q$ such that the q sums $\sum_{i \in I_j} i^k$ (with $j = 1, 2, \ldots, q$ and $k = 0, 1, \ldots, N - 1$) do not depend on j. He gave the following solution (for a proof see e.g., [42]): for each $q \geq 2$ define the sequence $\mathbf{T}_q = (T_q(n))_{n \geq 0}$ by $T_q(n) = s_q(n) \bmod q$. Then let $I_j = \{0 \leq i \leq q^N - 1 \ : \ T_q(i) = j\}$.

For an occurrence of related sequences, see [59]. A very nice relationship between magic cubes, Prouhet sequences, and the Prouhet-Tarry-Escott problem was given by Adler and Li [2]. For the state of art about the Prouhet-Tarry-Escott problem the reader can look at one of the surveys [24, 60].

5.2 Curious infinite products

Woods asked [68], what is the limit of the sequence

$$1/2, \quad \frac{1}{2}/\frac{3}{4}, \quad \frac{\frac{1}{2}}{\frac{3}{4}}/\frac{\frac{5}{6}}{\frac{7}{8}}, \quad \ldots \quad ?$$

Robbins [58] proved that this limit is $\frac{\sqrt{2}}{2}$. More precisely, we have the following:

Proposition 5. *Let $\varepsilon_n = (-1)^{t_n}$, where $(t_n)_{n\geq 0}$ is the Prouhet-Thue-Morse sequence. Then*

$$\left(\frac{1}{2}\right)^{\varepsilon_0}\left(\frac{3}{4}\right)^{\varepsilon_1}\left(\frac{5}{6}\right)^{\varepsilon_2}\cdots = \prod_{n=0}^{\infty}\left(\frac{2n+1}{2n+2}\right)^{\varepsilon_n} = \frac{\sqrt{2}}{2}. \tag{1}$$

Proof. We give a simple proof, discovered by the first author in 1987. Let P and Q be the infinite products defined by

$$P = \prod_{n=0}^{\infty}\left(\frac{2n+1}{2n+2}\right)^{\varepsilon_n}, \quad Q = \prod_{n=1}^{\infty}\left(\frac{2n}{2n+1}\right)^{\varepsilon_n}.$$

Then

$$PQ = \frac{1}{2}\prod_{n=1}^{\infty}\left(\frac{n}{n+1}\right)^{\varepsilon_n} = \frac{1}{2}\prod_{n=1}^{\infty}\left(\frac{2n}{2n+1}\right)^{\varepsilon_{2n}}\prod_{n=0}^{\infty}\left(\frac{2n+1}{2n+2}\right)^{\varepsilon_{2n+1}}.$$

Of course all products are convergent by Abel's theorem. Now, since $\varepsilon_{2n} = \varepsilon_n$, and $\varepsilon_{2n+1} = -\varepsilon_n$, we get

$$PQ = \frac{1}{2}\prod_{n=1}^{\infty}\left(\frac{2n}{2n+1}\right)^{\varepsilon_n}\left(\prod_{n=0}^{\infty}\left(\frac{2n+1}{2n+2}\right)^{\varepsilon_n}\right)^{-1} = \frac{1}{2}\frac{Q}{P}.$$

Since $Q \neq 0$, this gives $P^2 = 1/2$, and the result follows since P is positive. $\square$

Note that the mysterious number Q does not appear in the final result! No expression for Q in terms of known constants is currently known, nor is it known if Q is transcendental or even irrational. This number first occurred in a paper of Flajolet and Martin [36], who studied a class of probabilistic counting algorithms for estimating the number of distinct elements in a large collection of data. Their asymptotic analysis involves the constant $\varphi = .77351\cdots$ given by

$$\varphi = 2^{-1/2}e^{\gamma}\frac{2}{3}\prod_{n=1}^{\infty}\left(\frac{(4n+1)(4n+2)}{4n(4n+3)}\right)^{(-1)^{t_n}}$$

where γ is Euler's constant [36, Theorem 3.A]. It is clear that

$$Q = 2^{-1/2}e^{\gamma}\varphi^{-1}.$$

It is precisely while he was trying to compute Q (and hence φ) that the first author came across the proof above.

The infinite product (1) suggests trying to obtain the expansion

$$\left(\frac{1}{2}\right)^{\pm 1}\left(\frac{3}{4}\right)^{\pm 1}\left(\frac{5}{6}\right)^{\pm 1}\cdots$$

of a number β (for instance, $\frac{\sqrt{2}}{2}$) by a greedy algorithm, where the signs in the exponent are iteratively chosen so that the product thus far approximates β as closely as possible at each step. The following conjecture of the second author [63] was proved by the first author and H. Cohen [6].

Theorem 7 (Allouche, Cohen). *Define a sequence of signs $(\alpha_n)_{n\geq 0}$ by $\alpha_0 = 1$, and, if $\alpha_0, \alpha_1, \ldots, \alpha_n$ are known, define α_{n+1} by:*

$$\alpha_{n+1} = \begin{cases} +1 & \text{if } \left(\frac{1}{2}\right)^{\alpha_0}\left(\frac{3}{4}\right)^{\alpha_1}\cdots\left(\frac{2n+1}{2n+2}\right)^{\alpha_n} > \frac{\sqrt{2}}{2}, \\ -1 & \text{if } \left(\frac{1}{2}\right)^{\alpha_0}\left(\frac{3}{4}\right)^{\alpha_1}\cdots\left(\frac{2n+1}{2n+2}\right)^{\alpha_n} < \frac{\sqrt{2}}{2}. \end{cases}$$

Then the sequence $(\alpha_n)_{n\geq 0}$ is equal to the Prouhet-Thue-Morse sequence on the alphabet $\{-1, +1\}$, i.e., $\alpha_n = (-1)^{t_n}$ for all $n \geq 0$.

For generalizations of these results, see [63, 6, 13, 7].

5.3 Partitions of the set of integers

Let A be the (lexicographically) smallest set of integers such that 0 and 1 belong to A, and for each $n \geq 1$ that belongs to A, the number $2n$ does not belong to A. Hence the first few elements of A are

$$0\ \ 1\ \ 3\ \ 4\ \ 5\ \ 7\ \ 9\ \ 11\ \ 12\ \ 13\ \ 15\ \ 16\ \ 17\ \ 19\ \ 20\ \ 21\ \ 23\ \ \cdots$$

It is not hard to see that A and $2A = \{2x \ : \ x \in A\}$ form a partition of the set of non-negative integers; see [26, 5, 64]. An unexpected connection with the Thue-Morse sequence, proved in [5], is given below.

Theorem 8 (Allouche, Arnold, Berstel, Brlek, Jockusch, Plouffe, Sagan). *Let $(a_n)_{n\geq 0} = 0\ 1\ 3\ 4\ 5\ 7\ 9\ 11\ 12\ 13\ 15\ 16\ 17\ 19\ 20\ 21\ 23\ \cdots$ be the increasing sequence of the elements of A. Define the sequence $\mathbf{z} = (z_n)_{n\geq 0}$ by*

$$\mathbf{z} = 0^{a_1 - a_0}\ \ 1^{a_2 - a_1}\ \ 0^{a_3 - a_2}\ \ \cdots\ \ 0^{a_{2n+1} - a_{2n}}\ \ 1^{a_{2n+2} - a_{2n+1}} \cdots$$

where by c^j for $c \in \{0, 1\}$, we mean the string $\overbrace{cc\cdots c}^{j}$. In other words, the sequence of runs *of $\mathbf{z}$ is the first difference of the sequence $(a_n)_{n\geq 0}$. Then $\mathbf{z}$ is equal to the Prouhet-Thue-Morse sequence.*

5.4 Algebraicity of formal power series in positive characteristic

The Prouhet-Thue-Morse real number $\sum_{n\geq 0} t_n 2^{-n}$ was proved transcendental by Mahler [44, p. 363]; also see Dekking [32]. What can be said about the formal power series $\sum_{n\geq 0}^{\infty} t_n X^{-n}$? This series is transcendental over $\mathbb{Q}(X)$, as noted, for example, in [32]. But, considering this series modulo 2, we have the following proposition:

Proposition 6. *Let $F(X) = \sum_{n\geq 0} t_n X^{-n}$. Then F, considered as an element of $\mathbb{F}_2[[X^{-1}]]$, is quadratic over $\mathbb{F}_2(X)$. More precisely, we have*

$$(1+X)^3 F^2 + X(1+X)^2 F + X^2 = 0. \tag{2}$$

Proof. This is an easy consequence of the recurrence relations satisfied by the sequence $\mathbf{t}$. Namely, all computations being done modulo 2, we have

$$\begin{aligned}
F &= \sum_{n\geq 0} t_n X^{-n} = \sum_{n\geq 0} t_{2n} X^{-2n} + \sum_{n\geq 0} t_{2n+1} X^{-2n-1} \\
&= \sum_{n\geq 0} t_n X^{-2n} + X^{-1} \sum_{n\geq 0} (1+t_n) X^{-2n} \\
&= F^2 + X^{-1}\left(\frac{X^2}{1+X^2} + F^2\right) \\
&= \left(\frac{1+X}{X}\right) F^2 + \frac{X}{1+X^2} = \left(\frac{1+X}{X}\right) F^2 + \frac{X}{(1+X)^2}.
\end{aligned}$$

Hence, multiplying through by $X(1+X)^2$, we obtain Eq. (2). The fact that F is not a rational function is an easy consequence of the overlap-free property of the sequence $\mathbf{t}$.

More generally, a formal power series with coefficients in $\mathbb{F}_p$, where p is a prime number, is algebraic over the field $\mathbb{F}_p(X)$ if and only if the sequence of its coefficients is p-automatic. This theorem was proved by Christol [27], and more details are given in the paper of Christol, Kamae, Mendès France and Rauzy [28].

There is a theory of continued fractions for Laurent series with coefficients in a finite field [16]. In particular the continued fraction expansion of a quadratic series is ultimately periodic (see [47]; see [61] for a careful study when the ground field is not finite). The continued fraction expansion of $\sum_{n\geq 0} t_n X^{-n}$ is ultimately periodic with a pleasantly short period. It is given by

$$\sum_{n\geq 0} t_n X^{-n} = [0,\ X+1,\ \overline{X,\ X,\ X^3+X,\ X}\ ]$$

where, as usual, the vinculum denotes the repeating portion of the ultimately periodic continued fraction.

5.5 β-Expansions

Representing real numbers in non-integer bases goes back to Rényi [56] and Parry [54]. These expansions — sometimes called β-expansions — differ in some respects from the usual base-k expansions where k is an integer. For example, some numbers may have multiple representations. However, Komornik and Loreti recently proved the following theorem [41]:

Theorem 9 (Komornik, Loreti). *There exists a smallest real number β in the interval $(1,2)$, for which there exists a unique β-expansion of 1 in the form $1 = \sum_{n=1}^{\infty} \delta_n \beta^{-n}$, with $\delta_n \in \{0,1\}$. Furthermore, for this smallest β, the sequence of "digits" $(\delta_n)_{n\geq 1}$ satisfies $\delta_n = t_n$ for $n \geq 1$, where $\mathbf{t} = t_0 t_1 t_2 \cdots$ is the Prouhet-Thue-Morse sequence. The number β is the unique positive root of the equation $1 = \sum_{n=1}^{\infty} t_n \beta^{-n}$, and we have $\beta \doteq 1.787231650$.*

Komornik and Loreti proved [41] that the above result is a consequence of the following proposition:

Proposition 7. *The lexicographically least binary sequence $(w_n)_{n\geq 1}$ such that*

$$\begin{aligned} w_{n+1}\ w_{n+2} \cdots &< w_1\ w_2 \cdots \quad \text{if} \quad w_n = 0, \\ \overline{w_{n+1}}\ \overline{w_{n+2}} \cdots &< w_1\ w_2 \cdots \quad \text{if} \quad w_n = 1, \end{aligned}$$

(where the order is lexicographical order, and $\overline{0} = 1$, $\overline{1} = 0$) satisfies $w_n = t_n$ for $n \geq 1$.

The second author observed that this last result was previously stated in a slightly different form by the first author and M. Cosnard in [8]. See [9] and Section 7.1 below.

6 Semigroup and group theory

The Prouhet-Thue-Morse sequence $\mathbf{t}$ (or one of its variants) occurs in the solution of the Burnside problem for groups: *Is every group with a finite number of generators and satisfying the identity $x^n = 1$ finite?* The answer is yes (and well-known) if $n = 2$, since the group must be abelian in this case. But the answer is no for large odd n, since, as Novikov and Adian showed [52], an infinite group $\Gamma(m,n)$ on m generators and satisfying $x^n = 1$ for all $x \in \Gamma(m,n)$ exists for all $m > 1$, and for all odd n with $n \geq 4381$. Adian's book [1] gives more details about the result and its history, and improves the constant 4381 to 665. One of the steps in the proof consists of finding a cube-free binary sequence (see [1, p. 5]). Actually the cube-free binary sequence given there is *not* the Prouhet-Thue-Morse sequence, since it is not overlap-free. The author uses a result of Arshon [15] in order to construct a cube-free binary sequence, but in that paper Arshon actually gave a cube-free binary sequence that *is equal* to the Prouhet-Thue-Morse sequence on the alphabet $\{1,2\}$ (see [15, p. 779]).

One may also consider the Burnside problem for semigroups. As remarked by Morse and Hedlund [50, 57], with the aid of $\mathbf{v}$, the square-free sequence over $\{0,1,2\}$ given above, one can construct an infinite semigroup S on three generators such that $x^3 = x^2$ for all $x \in S$. Indeed, this is accomplished by letting $S = \{0,1,2\}^* \cup \{z\}$, where z is the zero element (i.e., $wz = zw = z$ for all $w \in S$) and subject to the relation $w^2 = z$ for all $w \in S$. Related questions were discussed by Brzozowski, Culik, and Gabrielian [25].

There is another occurrence of the Thue-Morse sequence in group theory, [22], as follows:

Theorem 10 (Boffa, Point). *Define the* Thue-Morse *group identities I_n by: $I_0(x,y)$ if $x = y$; $I_{n+1}(x,y)$ is $I_n(xy,yx)$. Then a finite group satisfies a Thue-Morse identity if and only if it is an extension of a nilpotent group by a 2-group.*

The reader will have noticed that

$$\begin{array}{l} I_0(x,y) \text{ if and only if } x = y \\ I_1(x,y) \text{ if and only if } xy = yx \\ I_2(x,y) \text{ if and only if } xyyx = yxxy \\ I_3(x,y) \text{ if and only if } xyyxyxxy = yxxyxyyx \\ \qquad\qquad \vdots \end{array}$$

and understood the terminology "Thue-Morse identities". See [23] for a generalization.

7 Real analysis

7.1 Iteration of continuous functions

Iterating a unimodal continuous function from $[0,1]$ into $[0,1]$ yields various behaviors going from convergent orbits to chaos; see the general reference [30]. M. Cosnard and the first author proved the following theorem [8] (also see [3, 31, 39]). Let f be a continuous map from $[0,1]$ into $[0,1]$, that is unimodal, i.e., increasing for x between 0 and some c, then decreasing for x between c and 1. Suppose that $f(1) = 0$. With the orbit $(f^{(n)}(1))_{n \geq 0}$ we associate the *itinerary* $(a_n)_{n \geq 0}$, defined by $a_n = 0$ if $0 \leq f^{(n)}(1) < c$ and $a_n = 1$ if $c < f^{(n)}(1) \leq 1$. With the sequence $\mathbf{a} = (a_n)_{n \geq 0}$ we associate the sequence $\widetilde{\mathbf{a}} = (\widetilde{a_n})_{n \geq 1}$ defined by

$$\widetilde{a_n} = \left(\sum_{j=0}^{n-1} a_j \right) \bmod 2.$$

Note that the sequence $\mathbf{a}$ is the first difference, taken modulo 2, of the sequence $\widetilde{\mathbf{a}}$.

Theorem 11 (Allouche, Cosnard). *The set of binary sequences $\widetilde{\mathbf{a}} = (\widetilde{a_n})_{n\geq 1}$ corresponding to unimodal continuous functions is exactly the set of binary sequences Γ defined by*

$$\Gamma = \{\mathbf{b} = (b_n)_{n\geq 1} : \ b_1 = b_2 = 1; \ \forall k, \ \overline{\mathbf{b}} \leq T^k(\mathbf{b}) \leq \mathbf{b}\},$$

where the order is lexicographical order, where $\overline{\mathbf{b}} = (\overline{b_n})_{n\geq 0} = (1 - b_n)_{n\geq 0}$, and where the map T is the shift defined by $T((w_n)_{n\geq 1}) = ((w_{n+1})_{n\geq 1})$. The least non-periodic element of Γ (which is also the least accumulation point of Γ) is $(t_n)_{n\geq 1}$ where $\mathbf{t} = t_0 t_1 t_2 \cdots$ is the Prouhet-Thue-Morse sequence.

Note that the classical approach does not use the transformation $\mathbf{a} \to \widetilde{\mathbf{a}}$; only the itineraries $\mathbf{a}$ are considered. The order between sequences is somewhat more complicated, although it boils down to the lexicographical order after applying the transformation $\mathbf{a} \to \widetilde{\mathbf{a}}$. The sequence $\mathbf{a}$ such that $\widetilde{\mathbf{a}} = (t_n)_{n\geq 1}$, i.e., the first difference of the Prouhet-Thue-Morse sequence, is

$$\mathbf{a} = 1 \ 0 \ 1 \ 1 \ 1 \ 0 \ 1 \ 0 \ \cdots$$

This sequence is called the *period-doubling sequence*. It is not hard to show that $\mathbf{a}$ is a fixed point of the morphism θ defined by $\theta(1) = 10$, $\theta(0) = 11$. For connections with Gray code, see [37].

The link between Theorem 11 and Proposition 7 above is easy [9]. Note also that Theorem 11 can be reformulated in number-theoretical terms:

Corollary 2. *Let Γ' the set of real numbers defined by*

$$\Gamma' = \{x \in [0,1] : \ \forall k \geq 0, \ 1 - x \leq \{2^k x\} \leq x\},$$

where $\{y\}$ denotes the fractional part of the real number y. Then the least irrational element of Γ' (which is also the least accumulation point of Γ') is the number $\tau = \sum_{n\geq 1} t_n 2^{-n} \doteq 0.824908$, where $\mathbf{t} = (t_n)_{n\geq 0}$ is the Prouhet-Thue-Morse sequence.

The number τ appears in other contexts. For example, let p be the probability that a randomly-chosen language L over $\{0,1\}$ has the property that there is at least one word of each possible length. (We flip a fair coin for each word w to decide if it is in L.) Then, as the second author has observed,

$$p = \prod_{i\geq 0}(1 - \frac{1}{2^{2^i}}) = \sum_{j\geq 0} \frac{(-1)^{t_j}}{2^j} = 2 - 2\tau.$$

7.2 The Knopp function

The Knopp function (see the introduction of [34]) is defined, for $a \in (0,1)$ and $b \in \mathbb{N} \setminus \{0\}$, by

$$K_{a,b}(x) = \sum_{n=0}^{\infty} a^n ||b^n x||$$

where $||y||$ is the distance from y to the nearest integer. In 1990, S. Dubuc and A. Elqortobi [34] came across the Prouhet-Thue-Morse sequence $\mathbf{t} = (t_n)_{n \geq 0}$ while studying the maximum of the Knopp function. They proved the following.

Theorem 12 (Dubuc, Elqortobi). *Let $a \in (0, 1)$ and let b be an even integer ≥ 2. Let $X^*(a, b)$ be the set of points where the function $K_{a,b}$ takes its maximum. Then the limit of the set $X^*(a, b)$, as $a \to (1/b)_-$, is the set $\{x, 1 - x\}$, where $x = ((b^2 - b)/2) \sum_{n=0}^{\infty} t_n/b^{n+1}$.*

8 Physics

Since the Prouhet-Thue-Morse sequence is both "easy to generate" and "non-trivial", it permits to generate a kind of *controlled disorder*. In particular this sequence has analogies (but also differences) with one-dimensional quasi-crystals: actually a typical one-dimensional quasicrystal is the Fibonacci sequence, i.e., the fixed point of the Fibonacci morphism $0 \to 01$, $1 \to 0$. Hence a large number of papers in physics study the Prouhet-Thue-Morse sequence. We only mention [17, 18], and the papers given in the bibliography of [12].

9 Generalizations

The alternative definitions of the Prouhet-Thue-Morse sequence given in Section 2 each suggest possible approaches to generalize the sequence.

For example, Proposition 1 suggests studying the generalized Prouhet-Thue-Morse sequence $\mathbf{t}_{k,m} = (s_k(n) \bmod m)_{n \geq 0}$ for integers $k \geq 2$ and $m \geq 1$. For example, we have

$$\mathbf{t}_{3,4} = 0\ 1\ 2\ 1\ 2\ 3\ 2\ 3\ 0\ 1\ 2\ 3\ \cdots.$$

Note that the sequence $\mathbf{t}_{2,m}$ has been studied by J. Tromp and the second author in [67], and that the sequence $\mathbf{t}_{q,q}$ is the sequence $\mathbf{T}_q$ of Section 5.1. Very recently the authors proved the following theorem [14], which generalizes the work of Thue:

Theorem 13. *Let $k \geq 2$, $m \geq 1$ be integers. The generalized Prouhet-Thue-Morse sequence $\mathbf{t}_{k,m}$ is overlap-free if and only if $m \geq k$.*

Several other generalizations of the Prouhet-Thue-Morse sequence have been studied, see for example [40, 45, 46, 53, 69].

10 Conclusion

The Prouhet-Thue-Morse sequence occurs in various fields, so that many apparently unrelated *definitions* of this sequence are equivalent. For example, Proposition 1, Theorem 2, Proposition 3, Theorem 4, Theorem 7, Theorem 8, Theorem 9, Proposition 7, Theorem 11, and even Theorem 10 or Corollary 2 can be turned into definitions. Automatic sequences, of which the Prouhet-Thue-Morse sequence is a simple example, are also useful because they are both "simple to generate" and "non-trivial": in physics as mentioned above, but also in other fields, such as music (see for example [11]). Searching for the many occurrences of the Prouhet-Thue-Morse sequence in the literature can be used as a pretext to take a delightful stroll through many fascinating areas of mathematics.

Acknowledgments

The first author wants to thank Professor Cunsheng Ding and Professor Kwok Yan Lam for having made it possible for him to come to Singapore for SETA98.

References

1. S. I. Adian, "The Burnside problem and identities in groups," Ergebnisse der Mathematik und ihrer Grenzgebiete **95**, Springer Verlag, 1979.
2. A. Adler and S.-Y. R. Li, Magic cubes and Prouhet sequences, *Amer. Math. Monthly* **84** (1977), 618–627.
3. J.-P. Allouche, "Théorie des Nombres et Automates," Thèse d'État, Université Bordeaux I, 1983.
4. J.-P. Allouche, Automates finis en théorie des nombres, *Exposition. Math.* **5** (1987), 239–266.
5. J.-P. Allouche, A. Arnold, J. Berstel, S. Brlek, W. Jockusch, S. Plouffe, and B. E. Sagan, A relative of the Thue-Morse sequence, *Discrete Math.* **139** (1995), 455–461.
6. J.-P. Allouche and H. Cohen, Dirichlet series and curious infinite products, *Bull. Lond. Math. Soc.* **17** (1985), 531–538.
7. J.-P. Allouche, H. Cohen, J. Shallit, and M. Mendès France, De nouveaux curieux produits infinis, *Acta Arith.* **49** (1987), 141–153.
8. J.-P. Allouche and M. Cosnard, Itérations de fonctions unimodales et suites engendrées par automates, *C. R. Acad. Sci. Paris Sér. I* **296** (1983), 159–162.
9. J.-P. Allouche and M. Cosnard, Non-integer bases, iteration of continuous real maps, and an arithmetic self-similar set, Preprint (submitted), 1998.
10. J.-P. Allouche, J. Currie, and J. Shallit, Extremal infinite overlap-free binary words, *Electronic J. Combinatorics* **5** (1) (1998), #R27.
11. J.-P. Allouche and T. Johnson, Finite automata and morphisms in assisted musical composition, *Journal of New Music Research* **24** (1995), 97–108.
12. J.-P. Allouche and M. Mendès France, Automata and automatic sequences, in "Beyond quasicrystals," F. Axel and D. Gratias, eds., Springer/Les Éditions de Physique, 1995, pp. 293–367.
13. J.-P. Allouche and J. Shallit, Infinite products associated with counting blocks in binary strings, *J. London Math. Soc.* **39** (1989), 193–204.
14. J.-P. Allouche and J. Shallit, Sums of digits and overlap-free words, in preparation.
15. C. E. Arshon, Proof of the existence of asymmetric infinite sequences, *Mat. Sb.* **44** (1937), 769–779.
16. E. Artin, Quadratische Körper im Gebiet der höheren Kongruenzen I, II, *Math. Zeitschrift* **19** (1924), 163–187. Reprinted in "Collected papers," pp. 1–104.
17. F. Axel, J.-P. Allouche, M. Kleman, M. Mendès France, and J. Peyrière, Vibrational modes in a one dimensional "quasi-alloy", *J. Physique*, Colloque C3, Supplement to No. 7 **47** (1986), C3-181–C3-186.

18. F. Axel and J. Peyrière, Spectrum and extended states in a harmonic chain with controlled disorder: effects of the Thue-Morse symmetry, *J. Statist. Phys.* **57** (1989), 1013–1047.
19. J. Berstel, Some recent results on squarefree words, in: "STACS84", M. Fontet and K. Mehlhorn, eds., Lecture Notes in Computer Science **166**, Springer Verlag, 1984, pp. 14–25.
20. J. Berstel, "Axel Thue's Papers on Repetitions in Words: a Translation," Publications du Laboratoire de Combinatoire et d'Informatique Mathématique **20**, Université du Québec à Montréal, 1995.
21. J. Berstel and P. Séébold, A characterization of overlap-free morphisms, *Disc. Appl. Math.* **46** (1993), 275–281.
22. M. Boffa and F. Point, Identités de Thue-Morse dans les groupes, *C. R. Acad. Sci. Paris Sér. I* **312** (1991), 667–670.
23. M. Boffa and F. Point, m-Identities, *C. R. Acad. Sci. Paris Sér. I* **314** (1991), 879–880.
24. P. Borwein and C. Ingalls, The Prouhet-Tarry-Escott problem revisited, *Enseign. Math.* **40** (1994), 3–27.
25. J. A. Brzozowski, K. Culik II, and A. Gabrielian, Classification of noncounting events, *J. Comput. System Sci.* **5** (1971), 41–53.
26. L. Carlitz, R. Scorville, and V. E. Hoggatt jr., Representations for a special sequence, *Fibonacci Quart.* **10** (1972), 499–518, 550.
27. G. Christol, Ensembles presque périodiques k-reconnaissables, *Theoret. Comput. Sci.* **9** (1979), 141–145.
28. G. Christol, T. Kamae, M. Mendès France, and G. Rauzy, Suites algébriques, automates et substitutions, *Bull. Soc. Math. France* **108** (1980), 401–419.
29. A. Cobham, Uniform tag sequences, *Math. Systems Theory* **6** (1972), 164–192.
30. P. Collet, J.-P. Eckmann, "Iterated maps on the interval as dynamical systems," Progress in Physics, Birkhäuser, 1980.
31. M. Cosnard, Étude de la classification topologique des fonctions unimodales, *Ann. Inst. Fourier* **35** (1985) 59–77.
32. M. Dekking, Transcendance du nombre de Thue-Morse, *C. R. Acad. Sci. Paris Sér. I* **285** (1977), 157–160.
33. F. M. Dekking, M. Mendès France, and A. van der Poorten, Folds!, *Math. Intelligencer* **4** (1982), 130–138, 173–181, 190–195.
34. S. Dubuc, A. Elqortobi, Le maximum de la fonction de Knopp, *Information Systems and Operational Research* **28** (1990), 311–323.
35. M. Euwe, Mengentheoretische Betrachtungen über das Schachspiel. *Proc. Konin. Akad. Wetenschappen, Amsterdam* **32** (1929), 633–642.
36. P. Flajolet and G. Nigel Martin, Probabilistic counting algorithms for data base applications, *J. Comput. Syst. Sci.* **31** (1985), 182–209.
37. H. Fredricksen, Gray codes and the Thue-Morse-Hedlund sequence, *J. Combin. Math. Combin. Comput.* **11** (1992), 3–11.
38. G. A. Hedlund, Remarks on the work of Axel Thue on sequences, *Nordisk Mat. Tidskrift* **15** (1967), 148–150.
39. L. Jonker, Periodic orbits and kneading invariants, *Proc. London Math. Soc.* **3** (1979), 428–450.
40. M. Keane, Generalized Morse sequences, *Z. Wahrscheinlichkeitstheorie Verw. Geb.* **10** (1968), 335–353.
41. V. Komornik and P. Loreti, Unique developments in non-integer bases, *Amer. Math. Monthly* **105** (1998), 636–639.

42. D. H. Lehmer, The Tarry-Escott problem, *Scripta Math.* **13** (1947), 37–41.
43. M. Lothaire, "Combinatorics on words," 2nd ed., Encyclopedia of Mathematics and its Applications **17**, 1997.
44. K. Mahler, Arithmetische Eigenschaften der Lösungen einer Klasse von Funktionalgleichungen, *Math. Annalen* **101** (1929), 342–266. Corrigendum, **103** (1930), 532.
45. J. C. Martin, Generalized Morse sequences on n symbols, *Proc. Amer. Math. Soc.* **54** (1976), 379–383.
46. J. C. Martin, The structure of generalized Morse minimal sets on n symbols, *Trans. Amer. Math. Soc.* **232** (1977), 343–355.
47. B. de Mathan, Approximations diophantiennes dans un corps local, *Bull. Soc. Math. France, Suppl. Mém.* **21** (1970).
48. M. Morse, Recurrent geodesics on a surface of negative curvature, *Trans. Amer. Math. Soc.* **22** (1921), 84–100.
49. M. Morse, Abstract 360: a solution of the problem of infinite play in chess, *Bull. Amer. Math. Soc.* **44** (1938), 632.
50. M. Morse and G. A. Hedlund, Symbolic dynamics, *Amer. J. Math.* **60** (1938), 815–866.
51. M. Morse and G. A. Hedlund, Unending chess, symbolic dynamics, and a problem in semigroups, *Duke Math. J.* **11** (1944), 1–7.
52. P. S. Novikov and S. I. Adian, Infinite periodic groups, I, II, III, *Izv. Akad. Nauk. SSSR Ser. Mat.* **32** (1968), 212–244, 251–524, 709–731.
53. R. Nürnberg, All generalized Morse-sequences are loosely Bernoulli, *Math. Zeitschrift* **182** (1983), 403–407.
54. W. Parry, On the β-expansions of real numbers, *Acta Math. Acad. Sci. Hung.* **11** (1960), 401–416.
55. E. Prouhet, Mémoire sur quelques relations entre les puissances des nombres, *C. R. Acad. Sci. Paris Sér. I* **33** (1851), 225.
56. A. Rényi, Representations for real numbers and their ergodic properties, *Acta Math. Acad. Sci. Hung.* **8** (1957), 477–493.
57. A. Restivo and C. Reutenauer, Rational languages and the Burside problem, *Theoret. Comput. Sci.* **40** (1985), 13–30.
58. D. Robbins, Solution to problem E 2692, *Amer. Math. Monthly* **86** (1979), 394–395.
59. J. B. Roberts, A curious sequence of signs, *Amer. Math. Monthly* **64** (1957), 317–322.
60. O. Salon, Le problème de Prouhet-Tarry-Escott, *Prétirage du LMD, Marseille* **94-23** (1994).
61. W. M. Schmidt, On continued fractions and Diophantine approximation in power series fields, Preprint, 1998.
62. P. Séébold, Sequences generated by infinitely iterated morphisms, *Disc. Appl. Math.* **11** (1985), 255–264.
63. J. Shallit, On infinite products associated with sums of digits, *J. Number Theory* **21** (1985), 128–134.
64. J.-i. Tamura, Partitions of the set of positive integers, nonperiodic sequences, and transcendence, in: "Analytic Number Theory," Kyoto, 1995, Sūrikaisekikenkyūsho Kōkyūroku, No. 961 (1996), pp. 161–182.
65. 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, pp. 139–158.

66. 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, pp. 413–478.
67. J. Tromp and J. Shallit, Subword complexity of a generalized Thue-Morse word, *Info. Proc. Letters* **54** (1995), 313–316.
68. D. R. Woods, Elementary problem proposal E 2692, *Amer. Math. Monthly* **85** (1978), 48.
69. J.-Y. Yao, Généralisations de la suite de Thue-Morse, *Ann. Sci. Math. Québec* **21** (1997), 177–189.