ELSEVIER

Discrete Mathematics 139 (1995) 455–461

DISCRETE MATHEMATICS

Note

A relative of the Thue–Morse sequence

Jean-Paul Allouche[a], André Arnold[b], Jean Berstel[c], Srečko Brlek[d], William Jockusch[e], Simon Plouffe[d], Bruce E. Sagan[f,*]

[a] C.N.R.S. U.R.A. 0226, Math., Université Bordeaux I, F-33405 Talence cedex, France
[b] LaBRI, Université Bordeaux I, F-33405 Talence cedex, France
[c] LITP, IBP, Université Paris VI, F-75252 Paris 05, France
[d] Département de Mathématiques et Informatique, UQAM, Montréal, Canada H3C 3P8
[e] Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA
[f] Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027, USA

Received 1 September 1992

Abstract

We study a sequence, c, which encodes the lengths of blocks in the Thue–Morse sequence. In particular, we show that the generating function for c is a simple product.

Keywords: Generating function; Thue–Morse sequence

Consider the sequence

$$\boldsymbol{c}\colon\ c_0, c_1, c_2, c_3, \ldots = 1, 3, 4, 5, 7, 9, 11, 12, 13, \ldots$$

defined to be the lexicographically least sequence of positive integers satisfying $n \in \boldsymbol{c}$ implies $2n \notin \boldsymbol{c}$. In fact, the lexicographic minimality of $\boldsymbol{c}$ makes it possible to replace the previous 'implies' with 'if and only if.' Equivalently, $\boldsymbol{c}$ is defined inductively by $c_0 = 1$ and

$$c_{k+1} = \begin{cases} c_k + 1 & \text{if } (c_k+1)/2 \notin \boldsymbol{c}, \\ c_k + 2 & \text{otherwise,} \end{cases} \tag{1}$$

for $k \geqslant 0$. This sequence was the focus of a problem of Kimberling in [7]. (In fact, he looked at the sequence $4c_0, 4c_1, 4c_2, \ldots$) The solution was given by Bloom [4]. Our Corollary 7 answers essentially the same question. Related results have recently been announced by Tamura [9].

* Corresponding author.

SSDI 0012-365X(93)00147-B

At the 4è Colloque Séries Formelles et Combinatoire Algébrique (Montréal, June 1992) Plouffe and Zimmermann [8] posed the following problem. Show that the generating function for $\boldsymbol{c}$ is

$$\sum_{k\geqslant 0} c_k x^k = \frac{1}{1-x}\prod_{j\geqslant 1}\frac{1-x^{2e_j}}{1-x^{e_j}} = \frac{1}{1-x}\prod_{j\geqslant 1}(1+x^{e_j}) \tag{2}$$

the sequence of exponents being

$$\boldsymbol{e}\colon\ e_1, e_2, e_3, e_4, \ldots = 1, 1, 3, 5, 11, 21, 43, \ldots,$$

where $e_1 = 1$ and

$$e_{j+1} = \begin{cases} 2e_j + 1 & \text{if } j \text{ is even,} \\ 2e_j - 1 & \text{if } j \text{ is odd,} \end{cases} \tag{3}$$

for $j \geqslant 1$. They found this conjecture by using a method that goes back to Euler. First they assumed that the generating function was of the form

$$\prod_{j\geqslant 0}\frac{1-x^{a_j}}{1-x^{b_j}}$$

for a certain pair of sequences a_j, b_j. Then they took the logarithm to convert the product into a sum. Finally they used Möbius inversion to determine the candidate sequences. Details of this procedure can be found in the text of Andrews [2, Theorem 10.3].

The purpose of this note is to prove (2). Before doing this, however, we will show that $\boldsymbol{c}$ has a number of other interesting properties. Chief among these is the fact that $\boldsymbol{c}$ is closely related to the famous Thue–Morse sequence, $\boldsymbol{t}$. See the survey article of Berstel [3] for more information about $\boldsymbol{t}$.

First we need to have a characterization of the integers in the sequence $\boldsymbol{c}$.

Proposition 1. *If n is any positive integer then $n \in \boldsymbol{c}$ if and only if $n = 2^{2i}(2j+1)$ for some nonnegative integers i and j.*

Proof. Every positive integer n can be uniquely written in the form $n = 2^k(2j+1)$ where $k, j \geqslant 0$. We will proceed by induction on k.

If $k = 0$, then n is odd. But then $n/2$ is not an integer, and so n is in the sequence by definition (1).

Now assume that $k \geqslant 1$ and that the proposition holds for all powers less than k of 2. If $k = 2i$ is even, then by induction we have $2^{2i-1}(2j+1) \notin \boldsymbol{c}$. So $n = 2^{2i}(2j+1) \in \boldsymbol{c}$ by (1). On the other hand, if $k = 2i+1$ is odd, then induction implies $2^{2i}(2j+1) \in \boldsymbol{c}$. Thus $n = 2^{2i+1}(2j+1) \notin \boldsymbol{c}$ as desired. $\square$

Let χ be the characteristic function of c, i.e.,

$$\chi(n)=\begin{cases}1 & \text{if } n\in c,\\ 0 & \text{otherwise.}\end{cases}$$

Restating the previous proposition in terms of χ yields the next result.

Lemma 2. *The function χ is uniquely determined by the equations*

$$\chi(2n+1)=1$$

$$\chi(4n+2)=0$$

$$\chi(4n)=\chi(n).$$

Another way of obtaining the sequence $\chi(n)$ for $n\geqslant 1$ is as follows. Starting from the sequence

$$101\bullet101\bullet101\bullet101\bullet\cdots$$

defined on the alphabet $\{0,1,\bullet\}$, fill in the successive holes with the successive terms of the sequence itself, obtaining:

$$101110101011101\bullet\cdots$$

Iterating this process infinitely many times (by inserting the initial sequence into the holes at each step), one gets a 'Toeplitz transform' which is nothing but our sequence χ. The proof of this fact is easily obtained using Lemma 2. See the article of Allouche and Bacher [1] for more information about Toeplitz transformations.

The connection with the Thue–Morse sequence can now be obtained. This sequence is

$$\boldsymbol{t}:\ t_0,t_1,t_2,t_3,\ldots=0,1,1,0,1,0,0,1,\ldots$$

defined by the conditions

$$t_0=0,$$

$$t_{2n+1}\equiv t_n+1 \pmod 2,$$

$$t_{2n}=t_n.$$

We will need a lemma relating $\boldsymbol{t}$ and χ. All congruences in this and any future results will be mod 2.

Lemma 3. *For every positive integer, n, we have*

$$\chi(n)\equiv t_n+t_{n-1}.$$

Proof. This is a three case induction based on Lemma 2 and the definitions of χ and $\boldsymbol{t}$. We will only do one of the cases as the others are similar.

$$\begin{aligned} t_{4n}+t_{4n-1} &\equiv t_{2n}+t_{2n-1}+1 \\ &\equiv t_n+t_{n-1}+2 \\ &\equiv \chi(n) \\ &= \chi(4n). \qquad \square \end{aligned}$$

Define d_k to be the first difference sequence of c_k, i.e., $d_k = c_k - c_{k-1}$, for $k \geqslant 0$ $(c_{-1}=0)$. So $\boldsymbol{d}$ is the sequence

$$d_0, d_1, d_2, d_3, d_4, \ldots = 1, 2, 1, 1, 2, 2, 2, 1, 1, 2, 1, \ldots$$

Note that from the definition of $\boldsymbol{c}$ in (1), the value of d_k is either 1 or 2. Write the Thue–Morse sequence in term of its blocks

$$\boldsymbol{t} = 011010011 \cdots = 0^{d'_0} 1^{d'_1} 0^{d'_2} 1^{d'_3} \cdots$$

defining a sequence d'_k. It is this sequence that is related to our original one via the difference operator.

Theorem 4. *For all $k \geqslant 0$ we have $d_k = d'_k$.*

Proof. Since both sequences consist of 1's and 2's, we need only verify that the 1's appear in the same places in both. It will be convenient to let $c'_k = \sum_{i \leqslant k} d'_i$. We now proceed by induction on k, assuming that $d_i = d'_i$ for $i \leqslant k$. Then, from the definitions,

$$d_{k+1} = 1 \iff \chi(c_k + 1) = 1. \tag{4}$$

But by the induction hypothesis, $c_k = \sum_{i \leqslant k} d_i = \sum_{i \leqslant k} d'_i = c'_k$. So, from Eq. (4),

$$\begin{aligned} d_{k+1} = 1 &\iff \chi(c'_k + 1) = 1 \\ &\iff t_{c'_k+1} + t_{c'_k} \equiv 1 \quad \text{(Lemma 3)} \\ &\iff t_{c'_k+1} \neq t_{c'_k} \\ &\iff d'_{k+1} = 1 \quad \text{(definitions).} \qquad \square \end{aligned}$$

Brlek [5] used the sequence $\boldsymbol{d}$ in calculating the number of factors of $\boldsymbol{t}$ of given length. The paper of de Luca and Varricchio [6] attacks the same problem in a different way.

Now if $n \in \boldsymbol{c}$ then we will consider its *rank*, $r(n)$, which is the function satisfying $c_{r(n)} = n$. Note that $r(n)$ is not defined for all positive integers n. In order to obtain a formula for $r(n)$, we will need a definition. Let the base 2 expansion of n be

$$n = \sum_{i \geqslant 0} \varepsilon_i 2^i$$

with the $\varepsilon_i \in \{0,1\}$ for all i. Define a function s by

$$s(n)=\sum_{i\geqslant 0}(-1)^i\varepsilon_i.$$

In other words, $s(n)$ is the alternating sum of the binary digits of n.

Theorem 5. *If $n\in c$ then*

$$r(n)=(2n+s(n))/3-1. \tag{5}$$

Proof. The proof will be by induction. From Proposition 1, $n\in c$ if and only if n is odd or $n=2^{2i}(2j+1)$ where $i>0$ and $j\geqslant 0$. To facilitate the induction, it will be convenient to split the odd numbers into two groups depending upon whether the highest power of 2 dividing $n+1$ is even or odd. So there will be three cases

(1) $n=2^{2i}(2j+1)$,
(2) $n=2^{2i}(2j+1)-1$,
(3) $n=2^{2i-1}(2j+1)-1$,

where $i>0$ and $j\geqslant 0$. The arguments are similar, so we will only do the first case.

So suppose n is even (remember that $i>0$). Thus $n+1$ is odd and, by Proposition 1, we have $n+1\in c$. Since both n and $n+1$ are in c, the left-hand side of Eq. (5) satisfies $r(n+1)=r(n)+1$. So, by induction, it suffices to show that $r'(n+1)=r'(n)+1$ where $r'(n)$ is the right-hand side of this equation. Moreover, n is a multiple of 4, hence $s(n+1)=s(n)+1$ (write down their binary expansions). Thus

$$\begin{aligned} r'(n+1)&=(2n+2+s(n+1))/3-1\\ &=(2n+2+s(n)+1)/3-1\\ &=(2n+s(n))/3\\ &=r'(n)+1. \qquad \square \end{aligned}$$

As a straightforward corollaries we have the next two results.

Corollary 6. *If $n\in c$ then*

$$r(n)=2n/3+\mathrm{O}(\log n)$$

and $r(n)$ takes the value $2n/3$ infinitely often.

Corollary 7. *For any nonnegative integer k*

$$c_k=3k/2+\mathrm{O}(\log k)$$

and $c_k=3k/2$ infinitely often.

We shall now prove the identity (2). First we note a property of the exponents e_j which is a simple consequence of their definition (3).

Lemma 8. *For $k \geqslant 2$, let $f_k = \sum_{2 \leqslant j \leqslant k} e_j$. Then*

$$f_k = \begin{cases} e_{k+1} - 2 & \text{if } k \text{ is even,} \\ e_{k+1} - 1 & \text{if } k \text{ is odd.} \end{cases}$$

Finally, we come to the proof. We restate the generating function here for easy reference.

Theorem 9. *The generating function for $\boldsymbol{c}$ is*

$$\sum_{k \geqslant 0} c_k x^k = \frac{1}{1-x} \prod_{j \geqslant 1} (1 + x^{e_j}).$$

Proof. It suffices to show that if $k \geqslant 2$ then

$$g_k(x) = \frac{1}{1-x}(1+x^1)(1+x^1)(1+x^3) \cdots (1+x^{e_k})$$

is the generating function for the sequence

$$1, 3, 4, 5, 7, \ldots, c_{f_k}, 2^k, 2^k, 2^k, \ldots$$

with $c_{f_k} = 2^k - 1$. The proof is an induction, breaking up into two parts depending on the parity of k. We will do the case where k is odd. (Even k is similar.) Now, by Lemma 8, $g_k(x)(1 + x^{e_{k+1}})$ is the generating function for the sequence

$$1, 3, \ldots, c_{f_k}, 2^k + 1, 2^k + 3, \ldots, 2^k + c_{f_k}, 2^{k+1}, 2^{k+1}, \ldots$$

Using Proposition 1 and the fact that k is odd, we see that $2^k + 1 = c_{f_k + 1}$ and $2^k + c_{f_k} = 2^{k+1} - 1 = c_{f_k + 1}$. So we want to show that

$$c_{f_k+1}, c_{f_k+2}, \ldots, c_{f_{k+1}} = 2^k + c_0, 2^k + c_1, \ldots, 2^k + c_{f_k}.$$

But if $n < 2^k$, then the highest power of 2 dividing n is equal to the highest power dividing $2^k + n$. Thus, by Proposition 1 again, $n \in \boldsymbol{c}$ if and only if $2^k + n \in \boldsymbol{c}$. This gives us the desired equality of the two sequences. □

One possible generalization of $\boldsymbol{c}$ is the sequence $\boldsymbol{c}^{(\alpha)}$ defined by $n \in \boldsymbol{c}^{(\alpha)}$ if and only if $\alpha n \notin \boldsymbol{c}^{(\alpha)}$. Thus $\boldsymbol{c}$ is the special case $\alpha = 2$.

The following observation is a direct consequence of our definitions.

Proposition 10. *If $\chi^{(\alpha)}(n)$ is the characteristic function of $c^{(\alpha)}$, then the sequence $(\chi^{(\alpha)}(n))$ is the unique fixed point of the morphism*

$$1 \to 1^{\alpha-1}0,$$

$$0 \to 1^{\alpha-1}1$$

which begins with 1.

One can also see that $c^{(\alpha)}$ satisfies analogs of many of our previous theorems. For example, if one defines $e_1^{(\alpha)} = 1$ and

$$e_{j+1}^{(\alpha)} = \begin{cases} \alpha e_j^{(\alpha)} + 1 & \text{if } j \text{ is even,} \\ \alpha e_j^{(\alpha)} - 1 & \text{if } j \text{ is odd,} \end{cases}$$

for $j \geqslant 1$, then the following result is a generalization of Theorem 9 and has an analogous proof.

Theorem 11. *The generating function for $c^{(\alpha)}$ is*

$$\frac{1}{1-x} \prod_{j \geqslant 1} \frac{1-x^{\alpha e_j^{(\alpha)}}}{1-x^{e_j^{(\alpha)}}}.$$

References

[1] J.-P. Allouche and R. Bacher, Toeplitz sequences, paperfolding, towers of Hanoi and progression-free sequences of integers, Ens. Math., 2e serie, 38, fascicule 3–4 (1992) 315–327.

[2] G. Andrews, q-series: their development and application in analysis, number theory, combinatorics, physics, and computer algebra, in: Conf. Board of the Math. Sci. Regional Conf. Ser. in Mathematics, Vol. 66 (American Mathematical Society, Providence, RI, 1986).

[3] J. Berstel, Axel Thue's work on repetitions in words, in: P. Leroux and C. Reutenauer, eds., Séries Formelles et Combinatoire Algébrique, Publications du LACIM, Vol. 11 (Université du Québec à Montréal, 1992) 65–80.

[4] D.M. Bloom, Solution to problem E 2850, Amer. Math. Monthly 89 (1982) 599–600.

[5] S. Brlek, Enumeration of factors in the Thue–Morse word, Discrete Appl. Math. 24 (1989) 83–96.

[6] A. de Luca and S. Varricchio, Some combinatorial properties of the Thue–Morse sequence, Theoret. Comput. Sci. 63 (1989) 333–348.

[7] C. Kimberling, Problem E 2850, Amer. Math. Monthly 87 (1980) 671.

[8] S. Plouffe and P. Zimmermann, Quelques conjectures, preprint (1992).

[9] J. Tamura, Some problems and results having their origin in the power series $\sum_{n=1}^{+\infty} z^{[\alpha n]}$ sum from 1 to infinity of z to the integral part of alpha times n, Reports of the Meeting on Analytic Theory of Numbers and Related Topics, Gakushuin Univ., 1992, 190–212.