Discrete Mathematics 43 (1983) 113–116
North-Holland Publishing Company

NOTE

NON-REPETITIVE SEQUENCES AND GRAY CODE

Helmut PRODINGER

Institut für Algebra und Diskrete Mathematik, Technische Universität Wien, Gußhausstraße 27–29, A-1040 Wien, Austria

Received 3 September 1981
Revised 5 February 1982

A sequence of 0's and 1's is constructed which is related to the Gray code, and which has only subwords ww of length not greater than ten.

1. Introduction

Consider a sequence $\omega = b_1 b_2 b_3 \cdots$, where $b_i \in \{0, 1\}$. A method to construct from this given sequence a new sequence $a_1 a_2 a_3 \cdots$ was proposed by Toeplitz (see Jacobs and Keane [2]):

The sequence $b_1 b_2 b_3 \cdots$ is written down, leaving a gap between every two symbols:

$$\begin{array}{ccccccc} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \cdots \\ b_1 & & b_2 & & b_3 & & b_4 \end{array}$$

Now the sequence $b_1 b_2 b_3 \cdots$ is filled into the gaps, leaving free every second gap. This last step is repeated *ad infinitum*, yielding the new sequence

$$T(\omega) = b_1 b_1 b_2 b_1 b_3 b_2 b_4 b_1 b_5 b_3 b_6 b_2 b_7 b_4 b_8 b_1 b_9 \cdots.$$

In [5] it is shown that $T(010101 \cdots)$ is a sequence of *bounded repetition*, i.e. only subwords ww of bounded length can occur. In particular, only subwords ww where the length of w is 1, 3 or 5 occur.

The sequence $010101 \cdots$ is in some sense the *base* of the *binary number system:* If $(n)_2 = s_m \cdots s_1 s_0$, the digits s_k form the sequence $0^{2^k} 1^{2^k} 0^{2^k} 1^{2^k} \cdots$ if n runs through the nonnegative integers.

There is another way to encode the integers by 0 and 1, the *Gray code.* A Gray code is an encoding of the integers as sequences of bits with the property that representations of adjacent integers differ in exactly one binary position. See [1, 4]. We restrict our considerations to the *standard Gray* (or *binary reflected*) code: If $(n)_{GR} = u_m \cdots u_1 u_0$ denotes the Gray code representation of n, then the digits u_k form the sequence $0^{2^k}1^{2^{k+1}}0^{2^{k+1}}1^{2^{k+1}}\cdots$ if n runs through the nonnegative integers. So one can consider the sequence $011001100\cdots$ as the basic sequence for the Gray code. In this note we are going to prove:

Theorem 1. *The sequence $a_1a_2a_3\cdots = 00101100\cdots$ obtained from the basic sequence of the Gray code by means of the construction of Toeplitz is of bounded repetition. In particular, only subwords ww where the length of w is 1, 2, 3 or 5 occur.*

As an example $a_{34}\cdots a_{38} = a_{39}\cdots a_{43} = 01011$.

2. Proof of Theorem 1

Let $p(n)$ be defined by $p(n)=1$ if $n\equiv 1 \pmod 4$ or $n\equiv 2 \pmod 4$ and $p(n)=0$ otherwise. Equivalently,

$$p(n)=\tfrac{1}{2}(1-(-1))^{\binom{n+1}{2}},$$

or, if $(n)_2 = u_m\cdots u_1u_0$, then $p(n)\equiv u_0+u_1 \pmod 2$. It is not hard to establish the following fact: If $(n)_2 = w10^l$ and w is the binary representation of m, then $a_n = p(m)$. The last two digits of $w=w'\sigma\tau$ determine $a_n : a_n \equiv \sigma+\tau \pmod 2$.

Since $a_2a_4a_6\cdots = a_1a_2a_3\cdots$, it is clear that if the subword ww with $|w|=n$ is impossible, then the subword ww with $|w|=2n$ is also impossible. So we prove that the subword ww is impossible for the length n of w:

(1) $n=4$; (2) $n=6, 10$; (3) $n=7$; (4) $n=9$; (5) $n=11$; (6) $n\geq 13$, n odd.

(1) Assume $a_{k+1}\cdots a_{k+4} = a_{k+5}\cdots a_{k+8}$ and let $i\in\{k+1, k+2\}$ be odd. Then $a_{i+4}=a_i$, which is impossible.

(2) Assume $a_{k+1}\cdots a_{k+6} = a_{k+7}\cdots a_{k+12}$ and let $i\in\{k+1, k+2\}$ be odd. Then $a_{i+6}=a_i$ and $a_{i+8}=a_{i+2}$; it is impossible that both equalities are fulfilled. For $n=10$ the argument is similar.

(3) If $a_{k+1}\cdots a_{k+7} = a_{k+8}\cdots a_{k+14}$ and $k=16m+i$, $0\leq i\leq 15$, a careful check of all 16 possibilities for i gives the proof.

(4) Similar as in (3), a check of all 32 possibilities for i modulo 32 gives the proof.

(5) The same argument as in (4) can be applied.

(6) Assume $a_{k+1}\cdots a_{k+n} = a_{k+n+1}\cdots a_{k+2n+1}$ and let $i\in\{k+1, k+2, k+3, k+4\}$ be the number with $i\equiv 2 \pmod 4$. Since $n+i$ is odd, we find that $a_ia_{i+2}a_{i+4}a_{i+6}a_{i+8}$ is either $abbaa$ or $aabba$ with $a\in\{0,1\}$. In both cases is $a_i = a_{i+8}$, which is impossible.

3. Further results

Let $n_1(k)$ be the number of 1's in $a_1\cdots a_k$. For the sequence $T(0101\cdots)$ the corresponding numbers have interesting properties according to the binary representation of k [5]. The same is true for the numbers $n_1(k)$.

First we give an estimate for the numbers $n_1(k)$.

Theorem 2. $n_1(k) = \frac{1}{2}k + O(\log k)$.

Proof. The sequence $b_1b_2b_3 \cdots = 01100 \cdots$ has the property that the number of ones in the first k places is $\frac{1}{2}k + O(1)$. The first k places of $a_1a_2a_3 \cdots$ only involve terms from $O(\log k)$ of the interleaved sequences, and each interleaved sequence can only contribute $O(1)$ to the error term.

Theorem 3.

$$n_1(k) = \sum_{i \geq 3} (\lfloor k/2^i + \tfrac{5}{8} \rfloor + \lfloor k/2^i + \tfrac{3}{8} \rfloor)$$

$$= \sum_{i \geq 2} \lfloor k/2^i + \tfrac{1}{4} \rfloor + \sum_{i \geq 3} (\lfloor k/2^i + \tfrac{3}{8} \rfloor - \lfloor k/2^i + \tfrac{1}{8} \rfloor).$$

Proof. Apply elementary counting arguments.

Theorem 4. $n_1(k) = \lfloor \frac{1}{4}k \rfloor + \lfloor \frac{1}{4}k + \frac{3}{4} \rfloor - B_2(1, k) + B_2(11, k) + B_2(101, k) + B_2(110, k)$ *where $B_2(w, k)$ denotes the number of occurrences of w as a subword of the binary representation of k with the convention that w is completed on the boundaries by zeroes (which is in this case important for $w = 110$).*

Proof.

$$n_1 = -\lfloor \tfrac{1}{2}k + \tfrac{1}{4} \rfloor + \sum_{i \geq 1} \lfloor k/2^i + \tfrac{1}{4} \rfloor - \lfloor \tfrac{1}{4}k + \tfrac{3}{8} \rfloor + \lfloor \tfrac{1}{4}k + \tfrac{1}{8} \rfloor - \lfloor \tfrac{1}{2}k + \tfrac{3}{8} \rfloor$$

$$+ \lfloor \tfrac{1}{2}k + \tfrac{1}{8} \rfloor + \sum_{i \geq 1} (\lfloor k/2^i + \tfrac{3}{8} \rfloor - \lfloor k/2^i + \tfrac{1}{4} \rfloor)$$

$$+ \sum_{i \geq 1} (\lfloor k/2^i + \tfrac{1}{4} \rfloor - \lfloor k/2^i + \tfrac{1}{8} \rfloor).$$

It is known [3, 6, 7] that the first sum equals $k - B_2(1, k) + B_2(11, k)$, that the second sum equals $B_2(101, k)$ and that the third sum equals $B_2(110, k)$. Furthermore

$$k - \lfloor \tfrac{1}{2}k + \tfrac{1}{4} \rfloor - \lfloor \tfrac{1}{4}k + \tfrac{3}{8} \rfloor + \lfloor \tfrac{1}{4}k + \tfrac{1}{8} \rfloor - \lfloor \tfrac{1}{2}k + \tfrac{3}{8} \rfloor + \lfloor \tfrac{1}{2}k + \tfrac{1}{8} \rfloor$$
$$= k - \lfloor \tfrac{1}{2}k \rfloor - \lfloor \tfrac{1}{4}k + \tfrac{1}{4} \rfloor + \lfloor \tfrac{1}{4}k \rfloor - \lfloor \tfrac{1}{2}k \rfloor + \lfloor \tfrac{1}{2}k \rfloor = \lfloor \tfrac{1}{4}k + \tfrac{3}{4} \rfloor + \lfloor \tfrac{1}{2}k \rfloor.$$

Remark. The Toeplitz construction scheme is, in some sense, a *binary scheme*. One could consider a Gray code scheme:

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	$a_{10} \cdots$
b_1			b_2	b_3			b_4	b_5	
	b_1					b_2			b_3
		b_1							
					b_1				

Each of the interleaved sequences acts as follows: take one, skip two, take two, skip two, take two, etc.

References

[1] P. Flajolet and L. Ramshaw, A note on Gray code and odd-even merge, SIAM J. Comput. 9 (1980) 142–158.

[2] K. Jacobs and M. Keane, 0–1-Sequences of Toeplitz type, Z. Wahrscheinlichkeitstheorie Verw. Gebiete 13 (1969) 123–131.

[3] P. Kirschenhofer, Subblock occurrences in the q-ary representation of n, preprint, SIAM J. Algebraic Discrete Methods (1983), to appear.

[4] D.E. Knuth, The Art of Computer Programming, Vol. 2, 2nd ed. (Addison-Wesley, Reading, MA, 1980).

[5] H. Prodinger and F.J. Urbanek, Infinite 0–1-sequences without long adjacent identical blocks, Discr. Math. 28 (1979) 277–289.

[6] H. Prodinger, Generalizing the sum of digits function, SIAM J. Algebraic Discrete Methods 3 (1982) 35–42.

[7] H. Prodinger, Subblock occurrences in representations of integers, preprint, TU Wien (1981).