Kolakoski Array over 2-letter Alphabet

N. Jansirani
Department of Mathematics, Queen Mary's College
Chennai 600 004, India. njansirani@gmail.com
V.R. Dare
Department of Mathematics, Madras Christian College
Chennai 600 059, India

Abstract

In this paper we introduce right angle path and layer of an array. We construct Kolakoski array and study some combinatorial properties of Kolakoski array. Also we obtain recurrence relation for layers and special elements.

Keywords: Complexity function, Kolakoski array, Layer, Special element and Ultimately Periodic.

1 Introduction

Combinatorics on words are of interesting importance in various fields of science like computer science, mathematics, biology, physics and crystallography. In this domain the combinatorial properties of infinite words and devices to generate these words are studied in literature [1, 3, 4]. Our motivation comes from formal language theory and therefore the combinatorial aspect will be stressed more than algebraic aspect. Formal languages with special combinatorial and structural properties are exploited in information processing or information transmission [5]. Moreover the theory has now developed into many directions and has generated a rapidly growing literature. In this paper we recall Kolakoski words over the 2-letter alphabet $\Sigma = \{1, 2\}$ which are invariant under the action of the run-length encoding operator. The widely known run-length encoding is used in many applications as a method for compressing data. For instance, the first step in the algorithm used for compressing the data transmitted by fax machines, consists of a run-length encoding of each line of pixels. In this paper we extended some special combinatorial one dimensional word properties to two dimensional arrays [5, 6].

An one sided infinite word z over the alphabet $\Sigma = \{1, 2\}$ is called a (classical) Kolakoski sequence, if it equals the sequence defined by its run length that is

$$\begin{array}{lllllllllllll} z = & \underline{22} & \underline{11} & \underline{2} & \underline{1} & \underline{22} & \underline{1} & \underline{22} & \underline{11} & \underline{2} & \underline{11\ldots} & & \\ & 2 & 2 & 1 & 1 & 2 & 1 & 2 & 2 & 1 & 2\ldots & = z \end{array}$$

Here, a run is a maximal sub word consisting of identical letters. The sequence $z' = 1z$,

$$\begin{array}{lllllllllllll} z' = & \underline{1} & \underline{22} & \underline{11} & \underline{2} & \underline{1} & \underline{22} & \underline{11} & \underline{2} & \underline{11} & \underline{22\ldots} & & \\ & 1 & 2 & 2 & 1 & 1 & 2 & 2 & 1 & 2 & 2\ldots & = z' \end{array}$$

is the only other sequence which has this property. The paper is structured as follows. In section 2 we recall some basic definitions and notations. In section 3 we define Kolakoski array and we give some combinatorial properties related to Kolakoski array. In addition to that we characterize the array with C^∞-words. In section 4 we present recurrence relation for special elements.

2 Basic Definitions

Let Σ be a finite alphabet. The set of all words over Σ is denoted by Σ^*. The empty word is denoted by λ. Let $\Sigma^+ = \Sigma^* - \{\lambda\}$. An infinite word w over a finite alphabet Σ is a mapping from positive integers into Σ. We write $w = a_1 a_2 \ldots a_i \ldots$ where $a_i \in \Sigma$. The set of all infinite words over Σ is denoted by Σ^ω. An infinite word w is ultimately periodic if $w = uv^\omega$. We will use the following convention
$r = min\{r, s\}$ and $s = max\{r, s\}$.

Let w be a word over $\Sigma = \{r, s\}$, $r, s \in N$. The set C^∞ contains the set of factors of any word w having the property that an arbitrary number of applications of the run-length encoding on w still produces a word over $\Sigma = \{r, s\}$. Such a word is called a smooth word. The definition of differentiable is chosen such that every sub word of a Kolakoski word is differentiable. In fact, every sub word of a Kolakoski word is smooth or a C^∞-word with respect to this differentiation rule over the respective alphabet, i.e., it is arbitrarily often differentiable [3, 4].

3 Kolakoski Array

In this section we introduce Kolakoski array and study the combinatorial properties of it. We start with some basic definitions which act as tools for the construction of the array. We also recall the definition of differentiable and smooth words [4]. Let $W_K \in \Sigma^{\omega\omega}$. The Kolakoski row (column) array is constructed by the rule:

.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
2	2	1	1	2	1	2	2	1	2	.	.
1	2	2	1	1	2	1	2	2	2	.	.
2	2	1	1	2	1	2	2	1	2	.	.
1	2	2	1	1	2	1	2	2	2	.	.
1	2	2	1	1	2	1	2	2	2	.	.
2	2	1	1	2	1	2	2	1	2	.	.
2	2	1	1	2	1	2	2	1	2	.	.

Figure 1: Kolakoski row array

W_K consists of consecutive blocks in every row (column) of 1's and 2's such that the length of each block is either 1 or 2, and the length of the i^{th} block in the j^{th} row (column) is equal $(i, j)^{th}$ vertex of W_K. These arrays are an example of a self reading infinite array.

.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
2	2	1	1	2	1	2	.	.	.	.	.
1	1	2	2	1	2	1	.	.	.	.	.
2	2	1	1	2	1	2	.	.	.	.	.
1	1	1	1	1	1	1	.	.	.	.	.
1	1	2	2	1	2	1	.	.	.	.	.
2	2	2	2	2	2	2	.	.	.	.	.
2	2	1	1	2	1	2	.	.	.	.	.

Figure 2: Kolakoski column array

Definition 3.1. Let $A = (a_{ij})$ be an infinite array. The element at $(i,i)^{th}$ position is denoted by a_i and is termed as i^{th} element of A.

Definition 3.2. Let $A = (a_{ij})$ be an infinite array. A word associated with a_{n1} to a_{1n} along $a_{n2}a_{n3}\dots a_{nn-1}a_{nn}a_{n-1n}\dots a_{2n}$ is called right angle word and the corresponding position in the array is called a right angle path.

Definition 3.3. Let $A = (a_{ij})$ be an infinite array. If the elements in a right angle path are equal then the path is called layer of width 1. If $r \in N$ consecutive right angle paths have same elements then it is treated as a single layer of width r.

Definition 3.4. For any $A \in \Sigma^{**}$, the sub array complexity of A is the map $g_A : N \times N \to N$ defined as $g_A(m,n) = card(S(A) \cap \Sigma^{m\times n})$ where $S(A)$ is the collection of all sub arrays of A. Then $g_A(m,n)$ counts the number of distinct sub arrays of A of size (m,n).

Definition 3.5. Let $W \in \Sigma^{\omega\omega}$ such that it consists of consecutive layers of 1's and 2's such that the width of each layer is either 1 or 2, and the width of the i^{th} layer is equal i^{th} special element of W. If we take 2 as the first element then odd layers consist of 2's and even ones of 1's. This construction produces self reading infinite arrays. These arrays are called Kolakoski arrays and one of the array is shown in the following figure.

.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
2	2	2	2	2	2	2	2	.	.	.	.
2	2	2	2	2	2	2	2	.	.	.	.
1	1	1	1	1	1	2	2	.	.	.	.
2	2	2	2	2	1	2	2	.	.	.	.
1	1	1	1	2	1	2	2	.	.	.	.
1	1	1	1	2	1	2	2	.	.	.	.
2	2	1	1	2	1	2	2	.	.	.	.
2	2	1	1	2	1	2	2	.	.	.	.

Figure 3: Kolakoski array

Definition 3.6. An array $A = (a_{ij})$ is said to be symmetric if $a_{ij} = a_{ji}$ for all $i, j \in N$.

Lemma 3.7. *Let $A = (a_{ij})$ be a Kolakoski array over a binary alphabet. Then A is a symmetric array.*

Proof. We note that both a_{ij} and a_{ji} lie in a layer and by definition of layer it follows that $A^T = A$.

Lemma 3.8. *Let $A = (a_{ij})$ be a Kolakoski array over a binary alphabet. Then $g_A(m,n) = g_A(n,m)$.*

Proof. Lemma 3.7 implies that the set of all sub arrays of size (m,n) becomes the set of all sub arrays of size (n,m). Hence $g_A(m,n) = g_A(n,m)$.

Lemma 3.9. *Let $A = (a_{ij})$ be a Kolakoski array over a binary alphabet $\Sigma = \{1,2\}$ with 2 as its first element. Then the word associated with $a_1a_2a_3\ldots$, is a famous Kolakoski word starting with 2.*

Proof. We get the proof from definition of layer and special element. We see that $a_1 = 2$; $a_2 = 2$; $a_3 = 1$; ... and so

$$\begin{array}{lcccccccccc} a_1a_2a_3a_4\cdots = & \underline{22} & \underline{11} & \underline{2} & \underline{1} & \underline{22} & \underline{1} & \underline{22} & \underline{11} & \underline{2} & \underline{11\ldots} \\ & 2 & 2 & 1 & 1 & 2 & 1 & 2 & 2 & 1 & 2\ldots \end{array}$$

which is a famous Kolakoski word starting with 2.

We are interested here in studying the set of factors of smooth words. For this, we recall some additional definitions. Let w be a word over Σ. Then w can be uniquely written as a concatenation of maximal blocks of identical symbols (called runs), i.e., $w = x^{i_1}x^{i_2}\ldots x^{i_n}$, with $x^{i_j} \in \Sigma$ and $i_j > 0$. The run-length encoding of w, noted $\Delta(w)$, is the sequence of exponents i_j, i.e., one has $\Delta(w) = i_1i_2\ldots i_n$. With respect to this differentiation rule over the respective alphabet we have the following definitions.

Definition 3.10. A word $w \in \Sigma$ is differentiable if $\Delta(w)$ is still a word over Σ.

Definition 3.11. An infinite word w over Σ is called a smooth word if for every integer $k > 0$ one has that $\Delta^k(w)$ is still a word over Σ.

Theorem 3.12. *Let $A = (a_{ij})$ be a Kolakoski array over a binary alphabet Σ. Then every word associated with special element of any sub array is a C^∞-word.*

Proof. We start the proof by considering a sub array $B = (b_{ij})$ of size (m,n), $m,n \in N$. Then the word associated with $b_1b_2b_3\ldots b_m$ $(m \le n)$ is a sub word of a Kolakoski word by Lemma 3.8. Since all sub words of a Kolakoski word are C^∞-words we infer that this result produces the statement that every diagonal word in a sub array of a Kolakoski array is a C^∞-word.

4 Recursive Formula

In this section we use the one dimensional recurrence properties to extend two dimensional [2]. Let $W_K \in \Sigma^{\omega\omega}$ be Kolakoski array. The n^{th} special element and is denoted by K_n.

We will now derive a recursive formula for K_n.

Let $k_n = min\left\{j : \sum_{i=1}^{j} K_i \geq n\right\}$.

Table 1: K_n and k_n

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
K_n	1	2	2	1	1	2	1	2	2	1	2	2	1	1	2
k_n	1	2	2	3	3	4	5	6	6	7	8	8	9	9	10

Lemma 4.1. $k_n = k_{n-1} + n - \sum_{i=1}^{k_{n-1}} K_i$, *where* $n \geq 2$.

Proof. We first notice that $n - 1 \leq \sum_{i=1}^{k_{n-1}} K_i \leq n$.
The left inequality holds by definition and the right one is valid, since if $\sum_{i=1}^{k_{n-1}} K_i \geq n + 1$ we would have $n - 1 \leq \sum_{i=1}^{k_{n-1}} K_i \geq n - 1$
which is a contradiction to the minimality of k_{n-1}. So, as the first case, we consider $\sum_{i=1}^{k_{n-1}} K_i = n - 1$ which implies $k_n = k_{n-1} + 1 = k_{n-1} + n - \sum_{i=1}^{k_{n-1}} K_i$.
In the second case $\sum_{i=1}^{k_{n-1}} K_i = n$ leads to $k_n = k_{n-1} = k_{n-1} + n - \sum_{i=1}^{k_{n-1}} K_i$.
We notice that Lemma 4.1 holds in general for every sequence, whose only values are 1 and 2.

Lemma 4.2.

$$k_n = k_{n-1} + |K_n - K_{n-1}| = 1 + \sum_{i=2}^{n} |K_i - K_{i-1}|, \quad \textit{where} \quad n \geq 2.$$

Proof. The following well known construction produces an array which is identical to Kolakoski array. Start with K_1 ones as special elements, continue with K_2 twos as special elements, followed by K_3 ones and so on. In this construction, after k_{n-1} steps two cases can appear, as described in the proof of Lemma 4.1. The first possibility is that $\sum_{i=1}^{k_{n-1}} K_i = n-1$ which means that we have constructed $n-1$ terms of the sequence. Therefore, by construction K_n must be different from K_{n-1} implying $k_n - k_{n-1} = |K_n - K_{n-1}|$. In the second case that $\sum_{i=1}^{k_{n-1}} K_i = n$, it is necessary that $K_{k_{n-1}} = 2$, for if otherwise $\sum_{i=1}^{k_{n-1}} K_i = n-1$, contradicting the miniality of k_{n-1}. So our construction has added 2 equal numbers at the k_{n-1}^{th} step, such that $K_n = K_{n-1}$ and finally $k_n - k_{n-1} = |K_n - K_{n-1}|$. The second equality follows by induction.

Corollary 4.3 is an implication of Lemma 4.2.

Corollary 4.3. $K_n \equiv$ (mod2 *or* $K_n = \frac{(-1)^{k_n}+1}{2} + 1$ *respectively.*

Corollary 4.4. $k_n = n - \frac{1}{2}\sum_{i=1}^{k_{n-1}}((-1)^{k_i}+1)$, *where* $n \geq 2$.

Corollary 4.4 follows from Corollary 4.3.

Corollary 4.5. $k_n = k_{n-1} + 1 - \frac{1}{2}(k_{n-1} - k_{n-2})((-1)^{k_{k_{n-1}}} + 1$, *where* $n \geq 3$.

Theorem 4.6. *For* $n \geq 3$, *we have*

$$K_n = K_{n-1} + (3 - 2K_{n-1})(n - \sum_{i=1}^{1+\sum_{j=2}^{n-1}|K_j - K_{j-1}|} K_i) \tag{1}$$

$$K_n = K_{n-1} + (3 - 2K_{n-1})(n - \sum_{i=1}^{1+\sum_{j=2}^{n-1}\frac{K_j - K_{j-1}}{3-2K_{j-1}}} K_i) \tag{2}$$

$$K_n = K_{n-1} + (3-2K_{n-1})\left(1 - \frac{1}{2}\frac{K_{n-1} - K_{n-2}}{3 - 2K_{n-2}}\left(1 + (-1)^{K^{1+\sum_{j=2}^{n-1}\frac{K_j - K_{j-1}}{3-2K_{j-1}}}}\right)\right) \tag{3}$$

Proof. From Lemma 4.1 and Lemma 4.2 we obtain

$$|K_n - K_{n-1}| = n - \sum_{i=1}^{1+\sum_{j=2}^{n-1}|K_j - K_{j-1}|} K_i$$

and use the fact that $|K_n - K_{n-1}| = \frac{K_n - K_{n-1}}{3-2K_{n-1}}$ to complete the proof of (1) and (2). The third equation (3) follows from Corollary 4.5 and Lemma 4.2.

Definition 4.7. We now define
$O_n = |\{1 \le j \le n : K_j = 1\}|$ and
$T_n = |\{1 \le j \le n : K_j = 2\}|$
The following table shows the first terms of the sequences defined above.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
K_n^1	1	2	2	1	1	2	1	2	2	1	2	2	1	1	2
K_n^2	2	2	1	1	2	1	2	2	1	2	2	1	1	2	1
O_n^1	1	1	1	8	17	17	30	30	30	49	49	49	74	101	101
T_n^1	0	3	8	8	8	19	19	34	51	51	72	95	95	95	124
O_n^2	0	0	5	12	12	23	23	23	40	40	40	73	98	98	127
T_n^2	1	4	4	4	13	13	26	41	41	60	81	81	81	108	108
$\|K_n\|$	1	3	5	7	9	11	13	15	17	19	21	23	25	27	29

From the above table we infer $O_{n+1} - O_n = |K_{n+1}|$ if $K_n = 1$ and $T_{n+1} - T_n = |K_{n+1}|$ if $K_n = 2$, $O_{n+1} = O_n$ if $K_n = 2$ and $T_{n+1} = T_n$ if $K_n = 1$

Lemma 4.8. *The following inequality is true for all positive integers n.*

$$n \le S_n < 2n \quad and \quad \frac{n}{2} < k_n \le n$$

where $S_n = \sum_{j=1}^{n} K_j$.

The following theorem is immediate from Lemma 4.8.

Theorem 4.9. *For all positive integers n, we have*

$$\frac{4n}{3} - 1 \le S_n \le \frac{5n}{3} \quad and \quad \frac{3n}{5} \le k_n \le \frac{3n}{4} + \frac{3}{2}$$

Theorem 4.10. *Kolakoski arrays are not ultimately periodic.*

Proof. The proof follows from a reason is that a (minimal) period of length q in a word of special elements z yields a period of length $q < q$ in its run-length word. Thus such a word z cannot be equal to its run-length word (where a word z is called ultimately periodic if there exist m, q such that $z_{i+1} \dots z_{i+q} = z_{i+q+1} \dots z_{i+2q}$ for all $i \ge m$).

Theorem 4.11. *Let $w \in \Sigma^{\omega\omega}$ with $||\Sigma|| \geq 2$. If g_ω is not bounded then $g_\omega(m,n) \geq m+n$ for all $n \geq 1$* [1].

Corollary 4.12. *Let $w \in \Sigma^{\omega\omega}$ with $||\Sigma|| \geq 2$. Then ω is ultimately periodic if and only if g_ω is bounded.*

Theorem 4.13. *Let $w \in \Sigma^{\omega\omega}$ be a Kolakoski array then $g_\omega(m,n) \geq m+n$ for all $m,n \geq 1$.*

Proof. The proof of the theorem follows from Theorem 4.10, Theorem 4.11, Corollary 4.12.

5 Conclusion

In this paper we defined a self reading infinite array over the 2-letter alphabet $\Sigma = \{1,2\}$ which is invariant under the action of the run-length encoding operator along the special element. Here we investigated interesting properties like recurrence formula, C^∞ sub array and ultimately periodicity of Kolakoski array and we seek more properties in this array which has many applications in the field of theoretical computer science and number theory.

References

[1] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*, Cambridge, England: Cambridge University Press (2003).

[2] Bernd Sing, *More Kolakoski Sequences*, arXiv:1009.4061v1, Cornell University, U.S (2010).

[3] Bertran Steinsky and Furbergstrasse, A Recursive Formula for the Kolakoski Sequence A000002, *Journal of Integer Sequences*, **9** (2006).

[4] S. Brlek, S. Duluq, A. Ladouceur and Vuillon, Combinatorial Properties of smooth infinite words, *Theor. Comput. Sci.*, **352** (2006), 306–317.

[5] F.M. Dekking, *On the structure of selfgenerating sequences*, Bordeaux (1980).

[6] C. Kimberling, Advanced Problem, *Amer. Math. Monthly.*, **86** (1979), 793.