TECHNICAL CONTRIBUTIONS

Repetitions in the Kolakovski sequence

ARTURO CARPI
Istituto di Cibernetica del C. N. R.
via Toiano 6, 80072 Arco Felice (NA), Italy
e-mail: arturo@arturo.na.cnr.it

The KOLAKOVSKI sequence [1] can be defined as the limit $\mathbf{k}$ of the sequence of words $\sigma^n(2)$ on the binary alphabet $\{1, 2\}$, where σ is the function inductively defined by

$$\sigma(\varepsilon) = \varepsilon,$$
$$\sigma(1x) = 2\,\overline{\sigma(x)}, \quad \sigma(2x) = 22\,\overline{\sigma(x)}, \quad \text{for all words } x,$$

ε and $\overline{w}$ denoting, respectively, the empty word and the word obtained by w by exchanging 1 and 2 (e. g., $\overline{211} = 122$). In other terms, $\mathbf{k}$ is the sequence on the alphabet $\{1, 2\}$ beginning by $\sigma^n(2)$, for all $n \geq 1$:

$$\mathbf{k} = 22112122122112\ldots.$$

As suggested by PĂUN [2], to exclude that a given word occurs in the KOLAKOVSKI sequence, one can proceed as follows: for any word x in which 111 and 222 do not occur, denote by $\varphi(x)$ the word obtained by deleting the eventual initial 1 and the eventual final 1 in the only word y such that $\sigma(y) \in \{x, \overline{x}\}$ (e.g., $\varphi(121122) = 122$). If 111 or 222 occur in x, or in $\varphi^n(x)$ for some $n \geq 1$, then x cannot occur in $\mathbf{k}$.

PĂUN used this method to prove that the KOLAKOVSKI sequence contains no '*square*' xx, with the length of x in the set

$$\{4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16\}$$

and no '*cube*' xxx, with x not longer than 16. He conjectured that the KOLAKOVSKI sequence is cube-free and that the length of the squares occurring in it is bounded. We shall give a positive answer to both conjectures.

A basic tool for our job is constituted by the following property of σ, which can be easily proved, by induction on the length of z.

Lemma 1. *If x, y, z are words such that $\sigma(z) = xy$ and the last letter of x differs from the first letter of y, then there exist words z_1, z_2 such that*

$$z = z_1 z_2, \quad \sigma(z_1) = x, \quad \sigma(z_2) \in \{y, \overline{y}\}.$$

In the sequel, as usual, the length of a word w will be denoted by $|w|$, and the words occurring in the KOLAKOVSKI sequence will be called its *factors*. Let S be the smallest set of words satisfying the following conditions:

S contains all factors of $\mathbf{k}$ not longer than 7,
$x \in S \Rightarrow \overline{x} \in S$,
$xy \in S \Rightarrow yx \in S$,
$x \in S$, $|x|$ even $\Rightarrow \sigma(x) \in S$.

Next lemma shows that S contains, in particular, the bases of the squares occurring in the KOLAKOVSKI sequence.

Lemma 2. *For any factor of $\mathbf{k}$ of the form xyx, with $|y| \leq 5$, one has $xy \in S$.*

Proof (outline): The statement is trivially true when $|xy| \leq 7$. Thus we suppose

$$\mathbf{k} = rxyx\mathbf{k}', \quad |xy| > 7 \tag{1}$$

for suitable r and $\mathbf{k}'$, and proceed by induction on $|xy|$.

One can always reduce himself to the case where the first letter of x is different from the last letter of xy and of r (provided r is non-empty), the last letter of x is different from the first letter of yx and of $\mathbf{k}'$, but the condition $|y| \leq 5$ is replaced by the weaker $|y| \leq 7$. Indeed, if, for instance, the first letter of x and the last letter of xy or of r are both a ($a \in \{1, 2\}$), then x begins by $a\overline{a}$, for $\mathbf{k}$ cannot contain aaa. Thus, deleting the first letter of x and replacing r and y respectively by ra and ya, one obtains that equations (1) are still satisfied, the first letter of x differs now from the last letter of xy and of r, and the new xy belongs to S if and only if the old one belongs, since it is obtained by 'moving' the initial a to the end of the word. In a symmetrical way, one can exclude that the last letter of x and the first letter of yx or of $\mathbf{k}'$ coincide.

Supposing that the previously quoted condition is verified, by Lemma 1 there is an initial factor of $\mathbf{k}$ of the form $r'zx'$ such that

$$\sigma(r') = r, \quad \sigma(z) \in \{xy, \overline{xy}\}, \quad \sigma(x') \in \{x, \overline{x}\},$$

The length of z is even, for the first and the last letter of $\sigma(z)$ are different. Still by Lemma 1, one has

$$z = x''y', \quad \sigma(x'') \in \{x, \overline{x}\} = \{\sigma(x'), \overline{\sigma(x')}\}, \quad \sigma(y') \in \{y, \overline{y}\},$$

for suitable words x'', y' and, consequently, $x'' = x'$. Thus $x'y'x'$ is a factor of $\mathbf{k}$. Moreover one has $|y'| \leq 5$ and $|x'y'| = |z| < |\sigma(z)| = |xy|$, because $|\sigma(v)| > 7 \geq |\sigma(y')|$ for all factor v of $\mathbf{k}$ of length 6, and $|u| < |\sigma(u)|$, for all factor u of $\mathbf{k}$, excepting ε, 1 and 11. For the sake of induction, we can then suppose $z = x'y' \in S$. One derives $\sigma(z) \in S$, and consequently $xy \in S$. □

The set S is finite, as one can verify by an exhaustive enumeration of its elements. More precisely the lengths of the words of S are:

1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 18, 27.

By Lemma 2, if a square xx occurs in $\mathbf{k}$, then one has $x \in S$. Thus, if $|x| \leq 16$, then $|x| \in \{1, 2, 3, 9\}$, as proved in [2], else $|x| \in \{18, 27\}$. Squares of words of length 27 actu-