Mathematical Notes, vol. 80, no. 5, 2006, pp. 710–714.
Translated from Matematicheskie Zametki, vol. 80, *no.* 5, 2006, *pp.* 751–756.

Finite-Automaton Transformations of Strictly Almost-Periodic Sequences

Yu. L. Pritykin

Received June 27, 2005

Abstract—Different versions of the notion of almost-periodicity are natural generalizations of the notion of periodicity. The notion of strict almost-periodicity appeared in symbolic dynamics, but later proved to be fruitful in mathematical logic and the theory of algorithms as well. In the paper, a class of essentially almost-periodic sequences (i.e., strictly almost-periodic sequences with an arbitrary prefix added at the beginning) is considered. It is proved that the property of essential almost-periodicity is preserved under finite-automaton transformations, as well as under the action of finite transducers. The class of essentially almost-periodic sequences is contained in the class of almost-periodic sequences. It is proved that this inclusion is strict.

Key words: *strictly almost-periodic sequence*, *finite automaton*, *finite transducer*.

1. INTRODUCTION

Strictly almost-periodic sequences (under a different name) were studied in the papers by Morse and Hedlund [1], [2] as well as by several other authors. This notion emerged in symbolic dynamics, but later proved to be fruitful both in mathematical logic and in the theory of algorithms.

The class of automaton images of strictly almost-periodic sequences (definitions are given below), obviously, contains the class of essentially almost-periodic sequences, by which we mean strictly almost-periodic sequences with an arbitrary prefix added at the beginning. Indeed, a sequence $a\omega$ can be obtained from a strictly almost periodic sequence ω by using what is called a finite time-lagged automaton: its memory stores a finite word a, and during its operation, it puts out this word first and then the input sequence with the time lag $|a|$ (to this end, the automaton stores the last $|a|$ symbols of the sequence in its memory). The main result of the paper (Theorem 2) consists in the fact that the two above-mentioned classes coincide. In other words, Theorem 2 claims that the property of essential almost-periodicity is preserved under finite-automaton transformations. In the last part of the paper, we consider an extension of the notion of a finite automaton, that of a finite transducer, and prove a similar statement for it.

It is worth mentioning that an extension of the class of strictly almost periodic sequences, that of almost-periodic sequences (see its definition below), was considered in [3]. In particular, it was proved that this class is also closed under automaton transformations. Clearly, the class of almost-periodic sequences contains the class of essentially strictly almost-periodic sequences. It turns out that this inclusion is strict (Theorem 1).

Let A be a finite alphabet. We shall consider sequences over this alphabet, i.e., maps $\omega\colon \mathbb{N} \to A$ (where $\mathbb{N} = \{0, 1, 2, \dots\}$). Denote by A^* the set of all strings over the alphabet A. If i and j, $i \le j$, are natural numbers, then we denote by $[i, j]$ the closed interval of the set of natural numbers with endpoints i and j, i.e., the set $\{i, i+1, i+2, \dots, j\}$. Denote by $\omega[i, j]$ the corresponding interval of the sequence ω, the string $\omega(i)\omega(i+1)\dots\omega(j)$. We say that $[i, j]$ is an *occurrence* of a string $x \in A^*$ in the sequence ω, if $\omega[i, j] = x$. Denote by $|x|$ the length of the string x. We think of a

 0001-4346/2006/8056-0710

sequence as written horizontally and extending from left to right infinitely; thus, speaking about smaller and greater indices, we shall say "to the left" and "to the right," respectively.

2. ALMOST-PERIODICITY

A sequence ω is called *almost-periodic* if for each string x that occurs in it infinitely many times, there exists a positive integer l such that each interval of length l of the sequence ω contains an occurrence of the string x.

A sequence ω is called *strictly almost-periodic* if for each string x that occurs in it at least once, there exists a positive integer l such that each interval of length l of the sequence ω contains an occurrence of the string x.

For convenience, let us introduce another definition. We shall say that a sequence ω is *essentially strictly almost-periodic* if it is the concatenation of a finite string and a strictly almost periodic sequence.

Each essentially strictly almost-periodic sequence is, obviously, almost periodic. Let us show that the class of almost-periodic sequences is strictly wider than the class of essentially strictly almost-periodic sequences.

Theorem 1. *There exists an almost-periodic sequence over the alphabet* $\{0,1\}$ *which is not essentially strictly almost-periodic.*

Proof. Let us construct a chain of binary strings:

$$a_0 = 1, \quad a_1 = 10011, \quad a_2 = 1001101100011001001110011,$$

and so on. The string a_{n+1} is obtained from a_n by the following rule:

$$a_{n+1} = a_n \overline{a}_n \overline{a}_n a_n a_n,$$

where $\overline{x}$ denotes the string obtained from x by replacing all ones by zeros and all zeros by ones. Set

$$c_n = \underbrace{a_n a_n \dots a_n}_{10}$$

and consider the sequence

$$\omega = c_0 c_1 c_2 c_3 \dots .$$

Let us prove that this is the desired sequence.

The length of a_n is 5^n; therefore, the length of the initial interval $c_0 c_1 \dots c_{n-1}$ of the sequence ω is equal to

$$10(1 + 5 + \dots + 5^{n-1}) = \frac{5}{2}(5^n - 1).$$

For convenience, let

$$l_n = \frac{5}{2}(5^n - 1) = |c_0 c_1 \dots c_{n-1}|.$$

Let us show that ω is almost-periodic. Suppose that a nonempty string x occurs in ω infinitely many times. Let us take a number n such that $|x| < 5^n$. Let $[i, j]$ be an occurrence of the string x in ω such that $i \geq l_n$. It follows from the construction of ω that for each k, the part of this sequence starting from the position l_k can be viewed not only as the concatenation of symbols 0 and 1, but also as the concatenation of the strings a_k and $\overline{a}_k$. Therefore, by the choice of i, the string x is a substring of one of the four strings $a_n a_n$, $a_n \overline{a}_n$, $\overline{a}_n a_n$, and $\overline{a}_n \overline{a}_n$. Notice that the string 10011 contains all the strings of length two (00, 01, 10, and 11). Therefore, the string a_{n+1} contains each of the strings $a_n a_n$, $a_n \overline{a}_n$, $\overline{a}_n a_n$, and $\overline{a}_n \overline{a}_n$. Hence x is a substring of a_{n+1}. Similarly, x is a substring of $\overline{a}_{n+1}$ (01100 also contains each of the strings 00, 01, 10, and 11). On each interval of length $2|a_{n+1}|$ to the right of the position l_{n+1}, there is an occurrence of a_{n+1} or $\overline{a}_{n+1}$; therefore, each interval of length $l = \frac{5}{2}(5^{n+1} - 1) + 2 \cdot 5^{n+1}$ includes an occurrence of the string x.

Now let us prove that for any positive integer n, the string c_n does not occur in the sequence ω to the right of the position l_{n+1}. It follows that for the sequence obtained from ω by cutting off the initial segment of a length not exceeding l_n, there exists a string, namely, c_n, which occurs in it a nonzero finite number of times. This will mean that this sequence is not strictly almost-periodic, and so ω is not an essentially strictly almost-periodic sequence.

Let ν be a sequence obtained from ω by cutting off the initial segment of length l_{n+1}. As we have already noticed, for each k, $1 \le k \le n+1$, it can be represented as the concatenation of the strings a_k and $\overline{a}_k$. Suppose that c_n occurs in ν, and let $[i, j]$ be one of its occurrences. For $n > 0$, the string c_n begins with a_1; therefore, $[i, i+4]$ is an occurrence of a_1 in ν. Notice that wherever $a_1 = 10011$ occurs in $a_1a_1 = 1001110011$, $a_1\overline{a}_1 = 1001101100$, $\overline{a}_1a_1 = 0110010011$, or $\overline{a}_1\overline{a}_1 = 0110001100$, it starts only from the zeroth or fifth position. Therefore, $5 \mid i$, i.e., the beginning of the occurrence of the string c_n in ν coincides with that of one of the strings a_1 or $\overline{a}_1$, which, as we can assume, constitute ν. Using induction on m, let us prove that $5^m \mid i$ for $1 \le m \le n$, i.e., if we represent ν as the concatenation of the strings a_m and $\overline{a}_m$, then the beginning of the occurrence of c_n coincides with the beginning of one of these strings. The base of induction (for $m = 1$) has already been proved. Assuming that this statement is true for $m = k$, we can view ν and c_n as being composed of the "characters" a_k and $\overline{a}_k$ with c_n occurring in ν. Then, to prove the statement for $m = k+1$, we can apply exactly the same argument as in the case $m = 1$, but with 1 and 0 replaced by a_k and $\overline{a}_k$, and a_1 and $\overline{a}_1$ replaced by a_{k+1} and $\overline{a}_{k+1}$; in so doing, we can use the fact that c_n begins with a_{k+1}.

Thus, we have shown that $5^n \mid i$; that is, if ν and c_n are viewed as composed of the "characters" a_n and $\overline{a}_n$, then we have proved that

$$c_n = \underbrace{a_n a_n \ldots a_n}_{10}$$

occurs in ν. But each interval of the sequence ν consisting of 10 successive "characters" a_n and $\overline{a}_n$ includes an occurrence of the "five-character" string a_{n+1} or $\overline{a}_{n+1}$, and this string contains an occurrence of the "character" $\overline{a}_n$. A contradiction. □

Moreover, we see from the proof how to construct an infinite, even a continuum, set of almost-periodic sequences that are not essentially strictly almost-periodic. For instance, we can proceed as follows: to each sequence $\tau\colon \mathbb{N} \to \{9, 10\}$, we can assign a sequence constructed the same way as ω from the proof of Theorem 1, but with the string

$$c_n^{(\tau)} = \underbrace{a_n a_n \ldots a_n}_{\tau(n)}$$

taken as c_n. Clearly, the ω_τ thus constructed will be different for different τ (for different τ_1 and τ_2, it suffices to consider the smallest n at which they differ: then $c_n^{(\tau_1)}$ and $c_n^{(\tau_2)}$ will contain different numbers of a_n, but they are followed in ω_{τ_1} and ω_{τ_2} by $a_{n+1} = a_n\overline{a}_n\overline{a}_n a_n a_n$). It remains to notice that there are continually many different τ.

3. AUTOMATON TRANSFORMATIONS

A *finite automaton* is a quintuple $F = \langle A, B, Q, \widetilde{q}, f \rangle$ consisting of finite sets A and B called input and output alphabets, respectively; a finite set of states Q; a designated set $\widetilde{q} \in Q$ called the initial state; and a transition function

$$f\colon Q \times A \to Q \times B.$$

A sequence $\langle p_n, \beta(n)\rangle_{n=0}^{\infty}$, where $p_n \in Q$, $\beta(n) \in B$, is called the *automaton transformation* of a sequence α of characters from the alphabet A if $p_0 = \widetilde{q}$ and $\langle p_{n+1}, \beta(n)\rangle = f(p_n, \alpha(n))$ for each n. The sequence β thus defined is denoted by $F(\alpha)$ and is called the *result* of the transformation of α by the automaton F. Obviously, for each F and α, the result $F(\alpha)$ exists and is uniquely defined. If $[i, j]$ is an occurrence of a string x in the sequence α with $p_i = q$, then we say that the automaton F *arrives* at this occurrence of x in the state q.

In [3], the following statement was proved: *if F is a finite automaton, and ω is an almost-periodic sequence, then $F(\omega)$ is also almost periodic.*

It turns out that this theorem can be supplemented.

Theorem 2. *If F is a finite automaton and ω is an essentially almost-periodic sequence, then $F(\omega)$ is also an essentially strictly almost-periodic sequence.*

Proof. Clearly, it will suffice to prove the theorem for strictly almost periodic sequences ω, because an essentially strictly almost periodic sequence (in our case, $F(\omega)$) remains essentially strictly almost-periodic after adding an arbitrary prefix to it.

Thus, let ω be a strictly almost-periodic sequence. By the theorem mentioned above, the sequence $F(\omega)$ is almost-periodic. Suppose that it is not essentially strictly almost-periodic. This means that for any positive integer N, there exists a string that occurs in $F(\omega)$ to the right of the position N and does not occur after that any more. Indeed, by cutting off the initial interval $[0, N]$ from $F(\omega)$, we shall not obtain a strictly almost-periodic sequence; therefore, there exists a string that occurs in it a nonzero finite number of times. Then we take the rightmost occurrence of this string.

Let $[i_0, j_0]$ be the rightmost occurrence of a certain string y_0 in $F(\omega)$. For a certain positive integer l_0, the string $x_0 = \omega[i_0, j_0]$ occurs in any interval of length l_0 of the sequence ω (by strict almost-periodicity of this sequence). Furthermore, if q_0 is the state in which the automaton F arrives at the position i_0, then it cannot arrive at any subsequent occurrence of the string x_0 in ω in the state q_0; otherwise it would put out the complete string y_0.

Now let $[r, s]$ be the rightmost occurrence of a certain string a in the sequence $F(\omega)$ with $r > i_0 + l_0$. There is an occurrence $[r', s']$ of the string x_0 on the interval $\omega[r - l_0, r]$, for which, by the choice of r, we have $r' > i_0$. Then we set

$$i_1 = r', \qquad j_1 = s, \qquad x_1 = \omega[i_1, j_1], \qquad y_1 = F(\omega)[i_1, j_1].$$

Since a does not occur in $F(\omega)$ to the right of the position r, the string y_1 that contains a as a substring, does not occur in $F(\omega)$ to the right of the position i_1. This means that if q_1 is the state in which the automaton arrives at the position i_1, then the automaton will never arrive at the occurrence of the string x_1 in ω in the state q_1 after that. Since x_1 begins with the string $\omega[r', s'] = x_0$ and $r' > i_0$, we have $q_1 \neq q_0$. Thus, we have found a string x_1 such that the automaton cannot arrive at any of its occurrences to the right of i_1 in the states q_0 or q_1.

Let $m = |Q|$ be the number of the automaton's states. Proceeding inductively along the same lines, we construct a chain of strings $x_k = \omega[i_k, j_k]$ and the corresponding distinct states q_k, where $k < m$, such that the automaton cannot arrive at any of the occurrences of x_k in ω to the right of i_k in the states $q_0, q_1, \dots, q_k$. For $k = m$, we obtain a contradiction. □

4. FINITE TRANSDUCERS

Let A and B be finite alphabets. A map $h\colon A^* \to B^*$ is called a *homomorphism* if for any $u, v \in A^*$ we have $h(uv) = h(u)h(v)$. Clearly, any homomorphism is completely defined by its values on single-character strings. If ω is a sequence of characters of the alphabet A, we set by definition

$$h(\omega) = h(\omega(0))h(\omega(1))h(\omega(2))\dots .$$

Let $h\colon A^* \to B^*$ be a homomorphism, and let ω be an almost-periodic sequence over the alphabet A. In [3], it was shown that the sequence $h(\omega)$ will also be almost-periodic. Then it is clear that if

ω is strictly almost-periodic, then $h(\omega)$ will also be strictly almost-periodic. Indeed, it suffices to prove that for any string $u = \omega[i, j]$, the string $h(u) = h(\omega(i)) \dots h(\omega(j))$ occurs in $h(\omega)$ infinitely many times. But this follows from the definition of $h(\omega)$ and from the fact that ω is strictly almost-periodic, which means that the string u occurs in it infinitely many times. Obviously, if ω is essentially almost-periodic, then $h(\omega)$ is also essentially almost-periodic.

A natural extension of the notion of finite automaton is that of a *finite transducer* (for more details, see [3], [4]). The difference is in that a finite transducer can produce a string of an arbitrary length having read one input symbol. Formally, only the definition of the transition function changes: now it is of the form $f\colon Q \times A \to Q \times B^*$. If a sequence $\langle p_n, v_n\rangle_{n=0}^{\infty}$, where $p_n \in Q$, $v_n \in B^*$, is a transformation of a sequence α, then the result of the transformation is the sequence $v_0 v_1 v_2 \dots$.

The action of a finite transducer can be represented as the composition of an automaton transformation and a homomorphism. Each of these types of transformations, as we know, preserves the property of almost-periodicity; thus, we obtain a corollary: almost-periodic sequences remain almost periodic under the action of finite transducers. Similarly, it follows from Theorem 2 and the preservation of essential strict almost periodicity under homomorphic maps that finite transducers also preserve the property of essential almost-periodicity.

ACKNOWLEDGMENTS

The author wishes to express his sincere gratitude to A. L. Semenov and A. A. Muchnik for drawing his attention to this topic as well as for valuable and useful remarks and advice concerning the paper.

This research was supported by the Russian Foundation for Basic Research under grant no. 06-01-00122 and by the program "Leading Scientific Schools" under grant no. 358.2003.1.

BIBLIOGRAPHY

1. M. Morse and G. A. Hedlund, "Symbolic dynamics," *Amer. J. Math.*, **60** (1938), 815–866.
2. M. Morse and G. A. Hedlund, "Symbolic dynamics II: Sturmian trajectories," *Amer. J. Math.*, **62** (1940), 1–42.
3. An. Muchnik, A. Semenov, and M. Ushakov, "Almost-periodic sequences," *Theoret. Comput. Sci.*, **304** (2003), 1–33.
4. A. Weber, "On the valuedness of finite transducers," *Acta Informatica*, **27** (1989), 749–780.

Yu. L. Pritykin
M. V. Lomonosov Moscow State University
E-mail: pritykin@lpcs.math.msu.su