Almost periodic sequences

An. Muchnik, A. Semenov, M. Ushakov

April 16, 2006

Abstract

This paper studies properties of almost periodic sequences (also known as uniformly recursive).

A sequence is almost periodic if for every finite string that occurs infinitely many times in the sequence there exists a number m such that every segment of length m contains an occurrence of the word.

We study closure properties of the set of almost periodic sequences, ways to generate such sequences (including a general way), computability issues and Kolmogorov complexity of prefixes of almost periodic sequences.

Keywords: almost periodic sequence, uniformly recurrent sequence, finite automaton, finite transducer, Kolmogorov complexity.

1 Introduction

Let Σ be a finite alphabet. We will talk of sequences in this alphabet, that is, functions from $\mathbb{N}$ to Σ (here $\mathbb{N} = \{0, 1, 2, \dots\}$).

Let $i, j \in \mathbb{N}$, $i \leq j$. Denote by $[i, j]$ the set $\{i, i+1, \dots, j\}$. Call this set a segment. If α is a sequence in an alphabet Σ and $[i, j]$ is a segment, then the string $\alpha(i)\alpha(i+1)\dots\alpha(j)$ is called a segment of α and written $\alpha[i, j]$. A segment $[i, j]$ is called an occurrence of a string u in a sequence α if $\alpha[i, j] = u$.

We imagine the sequences going horizontally from left to right, so we shall use terms "to the right" or "to the left" to talk about greater and smaller indices respectively.

Definition 1. A sequence $\alpha\colon \mathbb{N} \to \Sigma$ is called *almost periodic* if for any string u there exists a number m such that one of the following is true:

(1) There is no occurrence of u in α to the right of m.

(2) Any α's segment of length m contains at least one occurrence of u.

Let $\mathcal{AP}$ denote the class of all almost periodic sequences.

The notion of almost periodic sequences generalizes the notion of eventually periodic sequences (the sequence α is eventually periodic if there exist N and T such that $\alpha(n+T) = \alpha(n)$ for all $n > N$). We will prove further that there exists a continuum set of almost periodic sequences in a two-character alphabet (some examples of such continuum sets can be found in [4] and [12]). Obviously, the set of all eventually periodic sequences in any finite alphabet is countable.

Definition 2. A sequence $\alpha\colon \mathbb{N} \to \Sigma$ is called *strongly almost periodic* if for any string u either u does not have any occurrence in α or there exists a number m such that every segment of α of length m contains at least one occurrence of u.

Strongly almost periodic sequences (under a different name) were studied in the works of M. Morse, and G. A. Hedlund ([3], [4]). They have appeared first in the field of symbolic dynamics, but then turned out to be interesting in connection with computer science.

The notion of strong almost periodicity is not preserved even under the mappings given by the most simple algorithms, the finite automata. For example, a strongly almost periodic (and even periodic) sequence $0000\dots$ can be mapped by a finite automaton to a non-almost periodic sequence $1000\dots$. Finite automata map periodic sequences to eventually periodic, that is, becoming periodic after deleting some prefix. The property of eventual periodicity is preserved under the mappings done by finite automata. This leads to an idea to seek, for the notion of strong almost periodicity, a corresponding notion of eventual almost periodicity that would be preserved under the mappings done by finite automata. We succeeded at finding such a notion, and it is formulated in Definition 1. For brevity we called it simply *almost periodicity* (and not eventual almost periodicity).

The class of almost periodic sequences is significantly richer than the class of eventually periodic sequences and corresponds to a richer class of real-world situations. In many cases, however, studying bidirectional sequences (functions from $\mathbb{Z}$ to Σ) would be more adequate. We note that under a suitable definition the theory of bidirectional almost periodic sequences can be reduced to the theory of unidirectional almost periodic sequences, and study only unidirectional sequences.

This work studies the class $\mathcal{AP}$ in four directions. In Section 3 we study various closure properties of $\mathcal{AP}$. In Section 4 we consider methods of generating almost periodic sequences: block products (known from the paper [7]), dynamic systems (an example: the sign of $\sin(nx)$) and, finally, the universal method. In Section 5 we present some interesting examples of almost periodic sequences. Section 6 considers the Kolmogorov complexity of almost periodic sequences. The Section 2 is auxiliary; it presents some equivalent definitions of almost periodic sequences.

Some of this paper's results are a development of results published by one of the authors in [10].

2 Equivalent definitions

Consider all strings of length l. These are of two types: ones that occur in α only finitely many times and ones that have infinitely many occurrences. Let us call them type I and type II respectively. For any l there is a prefix of α such that it contains all occurrences of all strings of type I. Then, every string of length l occurring in the rest of α is of type II.

Consider a string u of type II. The above Definition 1 guarantees that gaps between occurrences of u in α are bounded above by some constant m. This fact can actually be taken as an equivalent definition of almost periodic sequences. By the "gap" between two occurrences $[i, j]$ and $[k, l]$ we understand $k - i$, the distance between the starting points of the occurrences.

Definition 3. A sequence α is almost periodic if for any l there exist numbers m and k such that every segment of length not more than l occurring to the right of k occurs infinitely many times in α and gaps between its occurrences are bounded above by m.

We stress that it is necessary to have m depend on l. The following theorem shows this:

Theorem 1. Let α be a sequence and m a number. Suppose that for every l there exists a number k such that every l-character segment of α to the right of k occurs infinitely many times in α and gaps between its occurrences do not exceed m. Then α is eventually periodic.

Proof. Let us show that α is eventually periodic and the period is at most $m!$. Consider k that corresponds to $l = m!$ in the statement of this theorem. We shall now prove that for every $i > k$, $\alpha(i) = \alpha(i + m!)$. Let i be greater than k and u be a string occurring in α in positions i through $i + m! - 1$. We are guaranteed that gaps between occurrences of u are no more than m. So, there is an occurrence of u starting at position j where $i < j \leq i + m - 1$. Since in that case $\alpha[i, i + m! - 1] = \alpha[j, j + m! - 1]$, we have

$$\alpha(i) = \alpha(j) = \alpha(i + (j - i)),$$
$$\alpha(i + (j - i)) = \alpha(j + (j - i)) = \alpha(i + 2(j - i)),$$
$$\dots$$

Taking into account that $j - i < m$ and thus $(j - i) \mid m!$, we get

$$\alpha(i) = \alpha(i + m!),$$

which proves the theorem. □

Finally, let us give an effective variant of our main definition.

Definition 4. An almost periodic sequence α is called *effectively almost periodic* if

- α is computable,
- m from Definition 1 is computable given u.

A parallel effective variant of Definition 3 is evidently equivalent to this one (we can take all strings of length $\leq l$ in turn, and choose maximal m; conversely, $m+k+l$ from the effective variant of Definition 3 fits any u of corresponding length l).

3 Closure properties of $\mathcal{AP}$

Denote by Σ^* the set of all strings in alphabet Σ including the empty string Λ.

Definition 5. A map $h\colon \Sigma^* \to \Delta^*$ is called a *homomorphism* if $h(uv) = h(u)h(v)$ for all $u, v \in \Sigma^*$. (We write uv for concatenation of u and v).

Clearly, homomorphism h is fully determined by its values on one-character strings. Let α be an infinite sequence of characters of Σ. By definition, put

$$h(\alpha) = h(\alpha(1))h(\alpha(2))\dots h(\alpha(n))\dots.$$

Evidently, if α is eventually periodic and $h(\alpha)$ is infinite, then $h(\alpha)$ is eventually periodic.

Theorem 2. Let $h\colon \Sigma^* \to \Delta^*$ be a homomorphism, and $\alpha\colon \mathbb{N} \to \Sigma$ be such a sequence that $h(\alpha)$ is infinite.

- If α is almost periodic, then so is $h(\alpha)$.
- If α is effectively almost periodic, then so is $h(\alpha)$.

Proof. Let us call a character $a \in \Sigma$ non-empty if $h(a) \neq \Lambda$. Since $h(\alpha)$ is infinite, there are infinitely many occurrences of non-empty characters in α. Now, since α is almost periodic, there exists a number k such that every α's segment of length k contains at least one non-empty character.

Take a natural number l. Every string of length l in $h(\alpha)$ is contained in the image of some string of length not more than kl in α.

Every single character in α maps into some segment of $h(\alpha)$ (which may be empty). Mark all ends of these segments for all characters of α. The sequence $h(\alpha)$ becomes separated into blocks of characters. All characters within such block map from a single character in α (and some blocks may be empty). Since Σ is finite, there exists an upper bound S on lengths of such blocks.

So, we found out that the homomorphism h can neither shrink nor expand the sequence "too much". The image of any segment of length L is no longer than LS and no shorter than $\frac{L}{k} - 1$. This is the main idea that leads us to the desired result. The following just fills in some technical details.

Let us take a prefix of α such that every string of length kl outside this prefix is of type II, and let m be a natural number bounding above the gaps between occurrences of these strings. Also let us take the corresponding prefix of $h(\alpha)$ and call $\tilde{h}$ the rest of $h(\alpha)$.

Consider an occurrence of any string u of length l in $\tilde{h}$. It is contained in the image of a string of length not more than kl. Let us denote this string by v and the corresponding segment of α by $[i, j]$. We have $|v| \leq kl$. By $\overline{v}$ denote the string of length kl in α starting at i. Every α's segment of length m contains a start of at least one occurrence of $\overline{v}$ in α. Let us prove that every $h(\alpha)$'s segment of length $(m+2)S$ contains a start of at least one occurrence of u.

Consider any segment of length $(m+2)S$ in $h(\alpha)$. It contains the image of an α's segment of length not less than $\frac{(m+2)S-2(S-1)}{S} \geq m$ (because every character in α maps to no more than S characters in $h(\alpha)$). This segment has a start of some occurrence of $\overline{v}$ in α. The image of this occurrence contains an occurrence of u in $h(\alpha)$. Therefore, the considered segment contains an occurrence of u.

To prove the second statement note that $h(\alpha)$ is computable and that $(m+2)S$ can be effectively computed. $\square$

Now let us study mappings done by finite automata.

Definition 6. A *finite automaton with output* is a tuple $\langle \Sigma, \Delta, Q, q_0, T \rangle$ where

- Σ is a finite set called *input alphabet*,
- Δ is a finite set called *output alphabet*,
- Q is a finite *set of states*,

- $q_0 \in Q$ is an *initial state*, and
- $T \subset Q \times \Sigma \times \Delta \times Q$ is a *transition set.*

If $\langle q, \sigma, \delta, q' \rangle \in T$, we say that the automaton in state q seeing the character σ goes to state q' and outputs the character δ.

Definition 7. If for any pair $\langle q, \sigma \rangle$ there exists a unique tuple $\langle q, \sigma, \delta, q' \rangle \in T$, the automaton is called *deterministic.*

Definition 8. Let α be a sequence and $\mathcal{A}$ an automaton. A sequence $(q_0, \delta_0), \ldots, (q_n, \delta_n), \ldots$ is a *run* of $\mathcal{A}$ on α if the following two conditions hold:

- q_0 is the initial state of $\mathcal{A}$, and
- $\langle q_i, \alpha(i), \delta_i, q_{i+1} \rangle$ is a transition of $\mathcal{A}$ for every $i \geq 0$.

Let us call $\delta_0, \ldots, \delta_n, \ldots$ the $\mathcal{A}$'s output on this run.

If $\mathcal{A}$ is deterministic, then it has a unique run on every sequence. Denote by $\mathcal{A}(\alpha)$ its output on α. (For an introduction in the theory of finite automata see, for example, [5].)

Theorem 3. Let $\mathcal{A}$ be a deterministic finite automaton and α an almost periodic sequence. Then $\mathcal{A}(\alpha)$ is also almost periodic. Moreover, if α is effectively almost periodic, then so is $\mathcal{A}(\alpha)$.

Proof. We need to prove that if some string u of length l occurs in $\mathcal{A}(\alpha)$ infinitely many times then the gaps between its occurrences are bounded above by a function in l. To prove this, it is sufficient to prove that for every occurrence $[i, j]$ of u located sufficiently far to the right in $\mathcal{A}(\alpha)$ there exists another occurrence of u within a bounded segment to the left of i. Obviously this already holds for α: there exist two monotone functions k and m such that for any l-character segment $[i, j]$ starting to the right of $k(l)$ there exists a "copy" of it starting between $i - m(l)$ and $i - 1$.

Take an l-character string $\widetilde{u}$ in $\mathcal{A}(\alpha)$ and its occurrence $[i, j]$. Suppose it is located sufficiently far to the right (leaving the exact meaning of "sufficiency" to a later discussion). Call u_1 the corresponding string in α (actually $u_1 = \alpha[i, j]$). Let $\mathcal{A}$ enter the segment $[i, j]$ in state q_1. For uniformity, denote $i_1 = i$ and $l_1 = l$.

There exists an occurrence of u_1 in α starting between $i_1 - m(l_1)$ and $i_1 - 1$. Denote the start of this occurrence i_2 and the corresponding $\mathcal{A}$'s state q_2. If $q_2 = q_1$ then $\mathcal{A}$ outputs the string $\widetilde{u}$ starting at i_2.

If $q_2 \neq q_1$ consider the string $u_2 = \alpha[i_2, j]$. Let l_2 be its length. This string has the following property. If $\mathcal{A}$ enters it in state q_1, it outputs $\widetilde{u}$ on the first segment of length l; if $\mathcal{A}$ enters it in state q_2, it enters the last segment of length l (which contains a copy of u_1) in state q_1 and, again, outputs $\widetilde{u}$.

There exists another occurrence of the string u_2 with a start between $i_2 - m(l_2)$ and $i_2 - 1$. Let i_3 be this start and q_3 the corresponding $\mathcal{A}$'s state.

If $q_3 = q_2$ or $q_3 = q_1$, then the automaton enters a copy of the string u_2 in state q_2 or q_1 and outputs $\widetilde{u}$ according to the formulated property. If $q_3 \neq q_2$ and $q_3 \neq q_1$, repeat the described procedure.

Namely, on the n'th step we have a string u_n of length l_n with an occurrence $[i_n, j]$ in α, and a set of states $q_1, \ldots, q_n$. The property is that if $\mathcal{A}$ enters u_n in one of the states $q_1, \ldots, q_n$, its output contains $\widetilde{u}$. Then, we find an occurrence of u_n with a start between $i_n - m(l_n)$ and $i_n - 1$, call its start i_{n+1} and the corresponding state q_{n+1}. If q_{n+1} equals one of the states $q_1, \ldots, q_n$, then we have found an occurrence of $\widetilde{u}$ to the left of i. Otherwise, we have found a string $u_{n+1} = \alpha[i_{n+1}, j]$ with a similar property. Since u_{n+1} starts with a copy of u_n, if $\mathcal{A}$ enters u_{n+1} in one of the states $q_1, \ldots, q_n$, it outputs $\widetilde{u}$ somewhere in this copy; if $\mathcal{A}$ enters u_{n+1} in state q_{n+1}, it outputs $\widetilde{u}$ at the end of u_{n+1}.

Since the set of $\mathcal{A}$'s states is finite, we only need to do the procedure a finite number of times, namely, $|Q|$ (here $|Q|$ is the cardinality of this set). After this number of steps we will definitely find another occurrence of $\widetilde{u}$.

Let us show that the gap between the found occurrence and the original occurrence $[i, j]$ is bounded from above. For the start of u_2 we have $i_2 \geq i_1 - m(l_1)$. Thus $l_2 \leq l_1 + m(l_1)$. To be able to take this step, we need $i_1 > k(l_1)$.

On the n'th step, we have

$$i_{n+1} \geq i_n - m(l_n) \geq i_1 - m(l_1) - m(l_2) - \ldots - m(l_n),$$

and

$$l_{n+1} \leq l_n + m(l_n) \leq l_1 + m(l_1) + m(l_2) + \ldots + m(l_n).$$

The n'th step can be performed if $i_n > k(l_n)$. To make this true, it is sufficient to have $i_1 - m(l_1) - \ldots - m(l_{n-1}) > k(l_n)$. So this is true if

$$\begin{array}{lcl} i_1 & > & k(l_1), \\ i_1 & > & k(l_2) + m(l_1), \\ i_1 & > & k(l_3) + m(l_1) + m(l_2), \\ \ldots & & \\ i_1 & > & k(l_{|Q|+1}) + m(l_1) + \ldots + m(l_{|Q|}). \end{array}$$

Let k' be the maximum of right-hand sides of these inequalities. Let $\widetilde{m} = l_{|Q|+1}$.

So, we proved that every string $\widetilde{u}$ that has an occurrence $[i, j]$ in $\mathcal{A}(\alpha)$ to the right of k' has another occurrence starting between $i - \widetilde{m}$ and $i - 1$. This suffices for $\mathcal{A}(\alpha)$ to be almost periodic. Our next goal is effectiveness issues.

Clearly, $\mathcal{A}(\alpha)$ is computable. If the sequence α is effectively almost periodic, then all mentioned numbers can be computed. We only need to be able to find out whether a given string $\widetilde{u}$ occurs in α finitely or infinitely many times.

To do so, consider a set S of all strings of length $\widetilde{m}$ that do not contain any occurrence of $\widetilde{u}$. There exist numbers k'' and $\widetilde{m}'$ such that every string in S that has an occurrence $[i', j']$ to the right of k'' has another occurrence starting between $i' - \widetilde{m}'$ and $i' - 1$. Let $K = \max\{k', k''\}$.

If there are infinitely many occurrences of $\widetilde{u}$, then every segment of length $\widetilde{m}$ has an occurrence of $\widetilde{u}$.

If, however, there are only finitely many occurrences of $\widetilde{u}$, then there is an occurrence of some string from S to the right of K. By shifting this occurrence to the left, we can find an occurrence with a start on the segment $[K, K + \widetilde{m}' - 1]$.

Note that if we found a segment of length $\widetilde{m}$ that does not have any occurrence of $\widetilde{u}$, then there is no occurrences to the right of it.

Now we can check the segment $[K, K + \widetilde{m}' + \widetilde{m} - 1]$ to see if it contains a subsegment of length $\widetilde{m}$ without an occurrence of $\widetilde{u}$. If we find such a subsegment, then there are finitely many occurrences of $\widetilde{u}$; otherwise, there are infinitely many occurrences. $\square$

Now we modify the definition of a finite automaton, allowing it to output any string (including the empty one) in the output alphabet when reading one character from input. These devices are usually called finite transducers. Formally, a transducer's transition set is a subset of $Q \times \Sigma \times \Delta^* \times Q$. The output sequence on the run $\langle q_0, v_0 \rangle, \ldots, \langle q_n, v_n \rangle, \ldots$ now is the concatenation $v_0 v_1 \ldots v_n \ldots$. (See [6].)

Define the program of effectively almost periodic sequence α to be a pair of two programs $\langle p_1, p_2 \rangle$ where p_1 is a program computing $\alpha(n)$ given n, and p_2 is a program computing m and k given l (as in Definition 3).

Corollary 4. Let $\mathcal{A}$ be a deterministic finite transducer with input alphabet Σ and output alphabet Δ, and $\alpha \colon \mathbb{N} \to \Sigma^*$ be a sequence such that the output sequence $\mathcal{A}(\alpha)$ is infinite. Then

1. if α is almost periodic, then so is $\mathcal{A}(\alpha)$, and
2. if α is effectively almost periodic, then $\mathcal{A}(\alpha)$ is effectively almost periodic, and the program for $\mathcal{A}(\alpha)$ can be effectively constructed given the program for α.

Proof. The proof follows from Theorems 2 and 3. We decompose the mapping done by the transducer into two: one will be a homomorphism and the other done by a finite automaton.

Define $f(\alpha)$ as follows: the i'th character of $f(\alpha)$ is a pair $\langle \alpha(i), q_i \rangle$, where q_i is the state of $\mathcal{A}$ when it reads the i'th character in α. Obviously, f can be done by a deterministic finite automaton. Then, define $g(\langle \sigma, q \rangle)$ as the string that $\mathcal{A}$ outputs when it reads σ in state q. Obviously, g is a homomorphism.

It is also clear that $g(f(\alpha)) = \mathcal{A}(\alpha)$. The effectiveness statement immediately follows from the mentioned theorems.

We also need to show that the programs for $\mathcal{A}(\alpha)$ can be effectively computed from the program for α. To do this, note that the proofs of Theorems 2 and 3 actually describe effective procedures. $\square$

Let α and β be two sequences $\alpha \colon \mathbb{N} \to \Sigma$ and $\beta \colon \mathbb{N} \to \Delta$. Define a cross product $\alpha \times \beta$ to be a sequence $\alpha \times \beta \colon \mathbb{N} \to \Sigma \times \Delta$ such that $(\alpha \times \beta)(i) = \langle \alpha(i), \beta(i) \rangle$.

We will show later that a cross product of two almost periodic sequences is not always almost periodic. On the other hand, a cross product of two eventually periodic sequences is eventually periodic.

Corollary 5. A cross product of an almost periodic sequence and an eventually periodic sequence is almost periodic.

Proof. The proof immediately follows from Theorem 3 since the cross product can be easily obtained as an output of a finite automaton reading the almost periodic sequence. □

Now we turn to nondeterministic transducers. Denote by $\mathcal{A}[\alpha]$ the set of all $\mathcal{A}$'s infinite output sequences on the input sequence α.

Theorem 6. (Theorem of uniformization.) Let $\mathcal{A}$ be a transducer and α an almost periodic sequence.

1. If $\mathcal{A}[\alpha] \neq \emptyset$ then there exists a deterministic transducer $\mathcal{B}$ such than $\mathcal{B}(\alpha) \in \mathcal{A}[\alpha]$ (so, $\mathcal{A}[\alpha]$ contains an almost periodic sequence).

2. If α is effectively almost periodic then given $\mathcal{A}$ and the program for α one can effectively compute if $\mathcal{A}[\alpha]$ is empty, and if it is not, effectively find $\mathcal{B}$.

Note that if α is not almost periodic then the uniformization could be impossible:

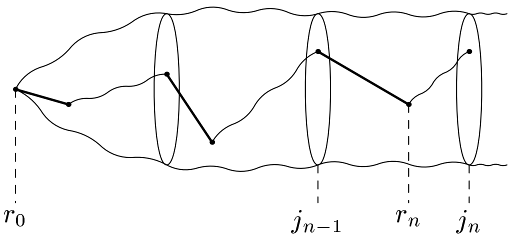
Let α be a sequence $\alpha = 01002000200000001\ldots$ (1s and 2s come in random order, and the number of separating zeroes increases infinitely). Let β be a sequence $\beta = 11222222211111111\ldots$ (every zero in a group is substituted by the character following that group). Then there exists a nondeterministic transducer $\mathcal{A}$ such that $\mathcal{A}[\alpha] = \{\beta\}$, but there is no deterministic transducer $\mathcal{B}$ such that $\mathcal{B}(\alpha) = \beta$.

Proof. Let us fix for the following the sequence α and introduce some terms. Any pair $\langle i, q\rangle$ where i is an integer and q is a state of $\mathcal{A}$, we call a point. We say that a point $\langle i_2, q_2\rangle$ is reachable from the point $\langle i_1, q_1\rangle$ if the transducer $\mathcal{A}$ can go from the state q_1 to the state q_2 reading $\alpha[i_1, i_2]$, namely, there exists a sequence

$$\langle s_{i_1}, u_{i_1}\rangle, \langle s_{i_1+1}, u_{i_1+1}\rangle, \ldots, \langle s_{i_2-1}, u_{i_2-1}\rangle, s_{i_2}$$

such that $s_{i_1} = q_1$, $s_{i_2} = q_2$, and for all $i \in [i_1, i_2 - 1]$ the tuple $\langle s_i, \alpha(i), u_i, s_{i+1}\rangle$ is a valid $\mathcal{A}$'s transition. The sequence $\langle s_{i_1}, u_{i_1}\rangle, \ldots, \langle s_{i_2-1}, u_{i_2-1}\rangle, s_{i_2}$ is called a path from $\langle i_1, q_1\rangle$ to $\langle i_2, q_2\rangle$, and the string $u_{i_1}u_{i_1+1}\ldots u_{i_2-1}$ is called the output string of this path. If there exists a path from $\langle i_1, q_1\rangle$ to $\langle i_2, q_2\rangle$ with a nonempty output string, we say that $\langle i_2, q_2\rangle$ is strongly reachable from $\langle i_1, q_1\rangle$. We say that a point is strongly reachable from a set of points if it is strongly reachable from some point in that set. Denote by $T_j(i, q)$ a set of points $\langle j, q'\rangle$ reachable from $\langle i, q\rangle$. Define $Q_j(i, q) = \{q' \mid \langle j, q'\rangle \in T_j(i, q)\}$.

Let $\langle r_0, s_0\rangle$ be some point. We say that a sequence $j_0 = r_0 < j_1 < \ldots < j_n < \ldots$ is correct with respect to $\langle r_0, s_0\rangle$ if for every $n \geq 1$ there exists a point $\langle r_n, s_n\rangle$ such that $j_{n-1} < r_n \leq j_n$, $\langle r_n, s_n\rangle$ is strongly reachable from $T_{j_{n-1}}(r_0, s_0)$, and $Q_{j_n}(r_0, s_0) = Q_{j_n}(r_n, s_n)$.



We sketch this on a figure. The dots represent points, the circle marked j_n represents $Q_{j_n}(r_n, s_n) = Q_{j_n}(r_0, s_0)$, the wavy lines in the center of the "tube" picture paths, and straight lines picture paths with a nonempty output string.

Say the point $\langle 0,$ the initial state of $\mathcal{A}\rangle$ is an initial point. A sequence is called correct if it is correct with respect to some point reachable from the initial point.

Introduce an equivalence relation "$\sim$" on a set of all points:

$$\langle i_1, q_1\rangle \sim \langle i_2, q_2\rangle \quad \text{iff} \quad \exists i \geq i_1, i_2 : Q_i(i_1, q_1) = Q_i(i_2, q_2).$$

This relation is obviously reflexive and symmetric. The transitivity property follows from the fact that if $Q_i(i_1, q_1) = Q_i(i_2, q_2)$ then for all $j > i$ $Q_j(i_1, q_1) = Q_j(i_2, q_2)$. This relation has another interesting property. If $\langle i_3, q_3\rangle$ is reachable from $\langle i_2, q_2\rangle$, $\langle i_2, q_2\rangle$ is reachable from $\langle i_1, q_1\rangle$, and $\langle i_1, q_1\rangle \sim \langle i_3, q_3\rangle$ then $\langle i_1, q_1\rangle \sim \langle i_2, q_2\rangle \sim \langle i_3, q_3\rangle$. This is so because for all $i \geq i_3$ we have $Q_i(i_3, q_3) \subset Q_i(i_2, q_2) \subset Q_i(i_1, q_1)$.

An amazing fact is that there can only be a finite set of equivalence classes, namely, not more than 2^N where N is the number of $\mathcal{A}$'s states. If there were 2^N+1 pairwise nonequivalent points $\{t_1,\ldots,t_{2^N+1}\}$ then for a sufficiently large i we would have 2^N+1 pairwise different sets $Q_i(t_1)$, $Q_i(t_2)$,..., $Q_i(t_{2^N+1})$, and that is impossible.

Now we are ready to prove the important

Lemma 7. $\mathcal{A}[\alpha]\neq\emptyset$ iff there exists a correct sequence.

Proof. If there is a correct sequence then surely $\mathcal{A}[\alpha]\neq\emptyset$: on the figure we see the path with a infinite output string drawn in the center of the "tube".

Now, suppose $\mathcal{A}[\alpha]\neq\emptyset$. Fix some run $\langle q_0,u_0\rangle,\ldots,\langle q_n,u_n\rangle,\ldots$ of $\mathcal{A}$ on α that has infinite output sequence $u_0u_1\ldots u_n\ldots$. Consider the sequence of points $\langle 0,q_0\rangle,\langle 1,q_1\rangle,\ldots,\langle n,q_n\rangle,\ldots$ where each point is reachable from the previous. Then these points separate into a finite set of equivalence classes:

$$\begin{gathered}\{\langle i,q_i\rangle \mid 0\le i\le i_1\},\\ \{\langle i,q_i\rangle \mid i_1< i\le i_2\},\\ \ldots\\ \{\langle i,q_i\rangle \mid i_m< i\}.\end{gathered}$$

We see that all points $\langle i,q_i\rangle$ where $i>i_m$ are equivalent. Now we can construct a correct sequence. Let $r_0=i_m+1$, $s_0=q_{r_0}$. We will construct two sequences j_n and $\langle r_n,s_n\rangle$ such that $j_{n-1}<r_n\le j_n$, $Q_{j_n}(r_n,s_n)=Q_{j_n}(r_0,s_0)$, and the point $\langle r_n,s_n\rangle$ is strongly reachable from $T_{j_{n-1}}(r_0,s_0)$. The state s_n will always be equal to q_{r_n}. Suppose we already found r_{n-1} and j_{n-1}. Let r_n be any number such that $r_n>j_{n-1}$ and the point $\langle r_n,q_{r_n}\rangle$ is strongly reachable from $T_{j_{n-1}}(r_0,s_0)$. We can find such a point because the output sequence of the path $\langle i,q_i\rangle$ is infinite. Since $\langle r_0,s_0\rangle\sim\langle r_n,q_{r_n}\rangle$, there exists a j_n such that $Q_{j_n}(r_n,q_{r_n})=Q_{j_n}(r_0,s_0)$. By induction, we now construct a correct sequence with respect to $\langle r_0,q_{r_0}\rangle$. Since that point is reachable from the initial point, we have constructed a correct sequence. The proof of the lemma is complete. $\square$

Lemma 8. (a) If α is almost periodic and $\mathcal{A}[\alpha]\neq\emptyset$ then there exists a correct sequence $j_0,j_1,\ldots,j_n,\ldots$ such that $\exists\mu\,\forall n\,(j_{n+1}-j_n)<\mu$.

(b) If α is effectively almost periodic then given $\mathcal{A}$ and the program for α one can find out if $\mathcal{A}[\alpha]$ is empty. If $\mathcal{A}[\alpha]\neq\emptyset$, one can find μ and a point $\langle r_0,s_0\rangle$ reachable from the initial point such that there exists a correct sequence j_n with $(j_{n+1}-j_n)<\mu$.

Proof. Let us construct an auxiliary deterministic finite automaton $\mathcal{C}$ with the output alphabet $\{0,1\}$. Among its states we will have a state $\bar{s}$ for every state s of $\mathcal{A}$.

We will need the following property of $\mathcal{C}$. Denote by $\mathcal{C}_{\langle r,s\rangle}(\alpha)$ the output sequence of $\mathcal{C}$ if we run it on α starting at time r in the state $\bar{s}$ (this sequence starts at index r; one can imagine its first r positions filled with zeroes). The property is that if there exists a correct sequence (for $\mathcal{A}$ and α) with respect to the point $\langle r,s\rangle$ then $\mathcal{C}_{\langle r,s\rangle}(\alpha)$ is a characteristic sequence of one such sequence. Otherwise, $\mathcal{C}_{\langle r,s\rangle}(\alpha)$ contains only a finite number of 1s. By characteristic sequence of a sequence $j_0<j_1<\ldots<j_n<\ldots$ we understand the sequence $\{a_i\}$ where

$$a_i=\begin{cases}1, & \text{if } \exists n\, i=j_n,\\ 0, & \text{otherwise.}\end{cases}$$

We describe the automaton $\mathcal{C}$ informally (omitting details regarding its states and transitions).

At the time r the automaton remembers s and print 1. At the time i $(i>r)$ the automaton remembers the following (we denote by j the last time less than i when $\mathcal{C}$ printed 1):

1. $Q_i(r,s)$,
2. the set of states $q\in Q_i(r,s)$ such that the point $\langle i,q\rangle$ is strongly reachable from $T_j(r,s)$, and
3. the class of all sets $Q_i(l,q)$ where $l\le i$ and the point $\langle l,q\rangle$ is strongly reachable from $T_j(r,s)$.

The automaton prints 1 if it sees that one of the sets from the third item equals to the set in the first item. Otherwise, it prints 0. It is obvious that the information remembered by the automaton is finite, and is bounded above by a function in the number of states of $\mathcal{A}$.

The needed property of $\mathcal{C}$ immediately follows from the fact that if there exists a correct sequence with respect to the point $\langle r,s\rangle$ then for all $i \geq r$ there exists a point that is strongly reachable from $T_i(r,s)$ and equivalent to $\langle r,s\rangle$.

Now we are ready to prove the statement (a) of the Lemma. Suppose $\mathcal{A}[\alpha] \neq \emptyset$. According to Lemma 7 there exists a correct sequence with respect to some point $\langle r_0,s_0\rangle$ reachable from the initial point. Then $\mathcal{C}_{\langle r_0,s_0\rangle}(\alpha)$ is a characteristic sequence of some correct sequence $j_0 < j_1 < \ldots$. If α is almost periodic then so is $C_{\langle r_0,s_0\rangle}(\alpha)$ according to Theorem 3. It follows that there exists μ such that $\forall n\,(j_{n+1}-j_n) < \mu$.

Now we turn to the statement (b). To prove it, we build another auxiliary deterministic finite automaton $\mathcal{D}$. We describe $\mathcal{D}$ informally, too. The idea is to find a point $\langle r,s\rangle$ such that there exists a correct sequence with respect to that point. To do this, the automaton $\mathcal{D}$ at time i runs a copy of the automaton $\mathcal{C}$ starting in every point $\langle i,s\rangle$ reachable from the initial point. It is impossible for a finite automaton to remember all these copies. But not all of these copies are different. Namely, at some time it can turn out that two copies are in the same state. Then these two copies are considered "united" and $\mathcal{D}$ may forget one of them. We will make it forget the one that was started later. So, at any time, $\mathcal{D}$ remembers a finite list of different states corresponding to remembered copies of $\mathcal{C}$. The later the copy was started the bigger its number in the list. Let $\mathcal{D}$ print a message "I am forgetting the copy number ν" when $\mathcal{D}$ forgets a copy. If some copy, say number ν, should print 1, let $\mathcal{D}$ print a message "The copy number ν prints 1". For convenience, let $\mathcal{D}$ print a message "I remember λ copies" every time.

If α is effectively almost periodic, then so is $\mathcal{D}(\alpha)$, so given $\mathcal{A}$ and the program for α we can compute the program for $\mathcal{D}(\alpha)$.

Every started copy will either be forgotten at some time or will survive infinitely. In the latter case its number in the list will stop decreasing sometime. Let γ be the number of such "survivors"; suppose they are started in points $t_1,\ldots,t_\gamma$. Let i_0 be the time when the numbers of "survivors" stop decreasing (and thus became equal $1,\ldots,\gamma$). Every later copy will eventually be forgotten, i.e. will unite with one of the "survivors". So, $\mathcal{A}[\alpha] \neq \emptyset$ iff one of the "survivors" prints infinitely many 1s. In other words, iff for some $\kappa \leq \gamma$ the automaton $\mathcal{D}$ prints infinitely many messages "The copy number κ prints 1".

If we know the program for $\mathcal{D}(\alpha)$, we can find the number γ (it is less by one than the smallest ν such that $\mathcal{D}$ prints "I am forgetting the copy number ν" infinitely many times), and know if there exists $i \leq \gamma$ with the required property. So, we can know whether $\mathcal{A}[\alpha] = \emptyset$. If $\mathcal{A}[\alpha] \neq \emptyset$, we can find i and the point t_i. Then there exists a correct sequence with respect to t_i and we can find μ (given a program for $\mathcal{D}(\alpha)$) such that the copy number i prints 1 on every segment of length μ, that is, there exists a correct sequence j_n such that for every n $(j_{n+1}-j_n) < \mu$. This completes the proof of the Lemma. □

Now we finish the proof of Theorem 6. Suppose $\mathcal{A}[\alpha] \neq \emptyset$ and α is almost periodic. We should build a deterministic finite transducer $\mathcal{B}$ for that $\mathcal{B}(\alpha) \in \mathcal{A}[\alpha]$. According to Lemma 8 we find a point $\langle r_0,s_0\rangle$ and a number μ such that there exists a correct (w.r.t. the point $\langle r_0,s_0\rangle$) sequence j_n such that for every n $(j_{n+1}-j_n) < \mu$. (When α is effectively almost periodic, this can be effectively found given $\mathcal{A}$ and the program for α).

Let $\mathcal{B}$ work as follows. Up to the time r_0 the transducer $\mathcal{B}$ prints an empty string. At the time r_0 the transducer prints an output string of any path from the initial point to the point $\langle r_0,s_0\rangle$. Then, $\mathcal{B}$ "marks" numbers j_n, r_n and states s_n such that

1. $j_{n-1} < r_n \leq j_n$,
2. $\langle r_n,s_n\rangle$ is strongly reachable from $T_{j_{n-1}}(r_0,s_0)$, and
3. $Q_{j_n}(r_n,s_n) = Q_{j_n}(r_0,s_0)$.

To do this, the transducer remembers at the time $i \geq r_0$ (here we denote by r and j the last positions marked as such):

1. $\alpha(i),\alpha(i-1),\ldots,\alpha(i-2\mu)$,
2. the last marked state s and a pair of numbers (μ_1,μ_2) such that $i-\mu_1 = j$ and $i-\mu_2 = r$,
3. $Q_{i-\mu_1}(r_0,s_0)$, $Q_i(r_0,s_0)$.

If $i - \mu_1 < i - \mu_2$, then the transducer searches for the next "j", so when it turns out that $Q_i(r_0, s_0) = Q_i(i - \mu_2, s)$, it marks i as the new "j". If $i - \mu_1 \geq i - \mu_2$, then the transducer searches for the next "r". To do this, it searches $T_i(r_0, s_0)$ for a point strongly reachable from $T_{i-\mu_1}(r_0, s_0)$, and, when it finds, marks the corresponding i as the new "r" and the corresponding state at the time i as the new "s". In this case, besides, the transducer prints the nonempty output string of some path from the last marked point $\langle r, s\rangle$ to the newly marked point. In all other cases $\mathcal{B}$ prints an empty string.

Since $j_n - r_{n-1} < 2\mu$, the remembered 2μ characters of α will suffice to know if the current i should be marked as "r" or "j", and to find the needed output string.

The output sequence of $\mathcal{B}$ is a concatenation of an infinite set of nonempty strings $u_0 u_1 \dots u_n \dots$ such that u_0 is an output string of a path from the initial point to $\langle r_0, s_0\rangle$, and for every $n > 0$ u_n is an output string of a path from $\langle r_{n-1}, s_{n-1}\rangle$ to $\langle r_n, s_n\rangle$. It follows that $\mathcal{B}(\alpha) \in \mathcal{A}[\alpha]$.

Since $\mathcal{B}$ can be effectively constructed, the proof is complete. □

4 Generating almost periodic sequences. The universal method

In the paper [7] an interesting method of generating infinite 0-1-sequences is presented. It is based on "block algebra".

4.1 Block product

Let u, v be strings in the alphabet $\{0, 1\}$ (we will use the symbol $\mathbb{B}$ for this alphabet from this point onwards, and also write $\mathbb{B}$-sequences in place of 0-1-sequences). The block product $u \otimes v$ is defined by induction on the length of v as follows:

$$u \otimes \Lambda = \Lambda$$
$$u \otimes v0 = (u \otimes v)u$$
$$u \otimes v1 = (u \otimes v)\bar{u},$$

where $\bar{u}$ is a string obtained from u by changing every 0 to 1 and vice versa. It is easy to check that block product is associative and right-distributive with respect to concatenation (that is, $u \otimes (v \otimes w) = (u \otimes v) \otimes w$, and $u \otimes (vw) = (u \otimes v)(u \otimes w)$, but not always $(uv) \otimes w = (u \otimes w)(v \otimes w)$).

Define the infinite block product. Let u_n, $n = 0, 1, \dots$ be a sequence of nonempty strings in the alphabet $\mathbb{B}$ such that for $n \geq 1$ u_n starts with 0. Then the product $\bigotimes_{n=0}^{\infty} u_n$ is defined as the limit of the sequence of strings u_0, $u_0 \otimes u_1$,…,$u_0 \otimes u_1 \dots \otimes u_n$,…. Since for every $n \geq 1$ u_n starts with 0, it follows that every string in this sequence is a prefix of the next string, so the sequence converges to some infinite $\mathbb{B}$-sequence.

In the paper [12] it is proved that for any sequence $\{u_n\}$ of strings that start with 0 and contain at least two characters their block product $\bigotimes_{n=0}^{\infty} u_n$ is strongly almost periodic. This fact allows us to prove that the cardinality of $\mathcal{AP}$ is continuum:

For a $\mathbb{B}$-sequence ω define $\alpha^\omega = \bigotimes_{n=0}^{\infty} (0\omega(n))$. Now the mapping $\omega \mapsto \alpha^\omega$ is an injection of continuum into $\mathcal{AP}$.

4.2 The universal method

Let Σ be a finite alphabet.

Definition 9. A sequence of tuples $\langle l_n, A_n, B_n\rangle$ where l_n is an increasing sequence of natural numbers, and A_n and B_n are non-empty finite sets of non-empty strings in the alphabet Σ, is called Σ-scheme if the following four conditions hold:

(C1) all strings in A_n have length l_n,

(C2) any string in B_n has the form $v_1 v_2$ where $v_1, v_2 \in A_n$, and every string from A_n is used as v_i in some string in B_n, (C3) every string u in A_{n+1} has the form $v_1 v_2 \dots v_k$ where for each $i < k$ $v_i v_{i+1} \in B_n$ (and thus $v_i, v_{i+1} \in A_n$) and for all $w \in B_n$ $\exists i < k\, w = v_i v_{i+1}$, and

(C4) every string u from B_{n+1} should have the following property: if $u = v_1 \dots v_k w_1 \dots w_k$ $(v_i, w_i \in A_n)$, then $v_k w_1 \in B_n$

Note that since all strings in A_n have equal lengths, the representation $u = v_1 \dots v_k$ of a string $u \in A_{n+1}$ is unique, and so is the representation $w = v_1 v_2$ of a string $w \in B_n$. Also note that $l_n \mid l_{n+1}$. A Σ-scheme is computable if the sequence $\langle l_n, A_n, B_n \rangle$ is computable.

Definition 10. We say that the sequence $\alpha \colon \mathbb{N} \to \Sigma$ is generated by a Σ-scheme $\langle l_n, A_n, B_n \rangle$ if for all $n \in \mathbb{N}$ there exists k_n such that for all $i \in \mathbb{N}$ $\alpha[k_n + il_n, k_n + (i+2)l_n - 1] \in B_n$, that is, a concatenation of any two successive strings in the sequence

$$\alpha[k_n, k_n + l_n - 1], \alpha[k_n + l_n, k_n + 2l_n - 1], \dots$$

is in B_n.

The sequence is perfectly generated by the scheme if $l_n \mid k_n$.

The sequence is effectively generated if the sequence k_n is computable.

Proposition 9. Any scheme perfectly generates some sequence.

Proof. Let $\langle l_n, A_n, B_n \rangle$ be any scheme. Consider an infinite tree of strings. Its nodes at n'th level are strings of length l_n, and the string x is the string's y parent if x is a prefix of y.

At n'th level mark the nodes x for which the following condition holds:

$$\forall i < n \forall j\, x[jl_i, (j+2)l_i - 1] \in B_i.$$

(I.e. the strings that can be prefixes of a sequence perfectly generated by the considered scheme.) Let us show that if some node is marked, then all its predecessors are marked, too. This follows, by induction, from the properties (C3) and (C4).

There are infinitely many marked nodes, because every string in A_n is marked. Hence, due to the compactness of Cantor space, there exists an infinite path in the tree with all its nodes marked. Consider a limit sequence of this path. It is perfectly generated by the scheme. □

Theorem 10. (a) Either of the next two properties of a sequence $\alpha \colon \mathbb{N} \to \Sigma$ is equivalent to the almost periodicity of α:

- α is generated by some Σ-scheme,
- α is perfectly generated by some Σ-scheme.

(b) Either of the next two properties of a computable sequence $\alpha \colon \mathbb{N} \to \Sigma$ is equivalent to the effective almost periodicity of α:

- α is effectively generated by some computable Σ-scheme,
- α is effectively and perfectly generated by some computable Σ-scheme.

Proof. We start with proving (a). Suppose α is generated by some Σ-scheme $\langle l_n, A_n, B_n \rangle$. Let us prove that α is almost periodic. Take a string $u \in \Sigma^*$ such that u has infinitely many occurrences in α. We prove that for some N every α's segment of length N has an occurrence of u. Denote the length of u by $|u|$. Take n such that $l_n \geq |u|$. Let us prove that every string in A_{n+1} contains u as a substring. Take k_n from the Definition 10. Since u has infinitely many occurrences in α, there exists an occurrence of u to the right of k_n, starting, say, on a segment $[k_n + il_n, k_n + (i+1)l_n - 1]$. Since $|u| \leq l_n$, the whole occurrence is contained in the segment $[k_n + il_n, k_n + (i+2)l_n - 1]$. According to the same Definition, this segment of α is in B_n. So, some string in B_n contains u. It follows that every string in A_{n+1} contains u since every string in A_{n+1} contains all strings from B_n (see (C3)).

Now, due to the definition of generation and to (C2), (C3), there exists k_{n+1} such that for every i

$$\alpha[k_{n+1} + il_{n+1}, k_{n+1} + (i+1)l_{n+1} - 1] \in A_{n+1},$$

and thus every α's segment of length $3l_{n+1}$ to the right of k_{n+1} contains at least one occurrence of some string from A_{n+1}, and thus, an occurrence of u.

Now suppose α is almost periodic. We construct a scheme $\langle l_n, A_n, B_n\rangle$ that perfectly generates α. Say that the occurrence $[i, i+|u|-1]$ of the string $u \in A_n \cup B_n$ in α is good if $l_n \mid i$. Let

$$A_n = \{u \in \Sigma^{l_n} \mid \ u \text{ has infinitely many good occurrences in } \alpha\},$$
$$B_n = \{u \in \Sigma^{2l_n} \mid \ u \text{ has infinitely many good occurrences in } \alpha\}.$$

We still need to define l_n. We do this by induction. Let $l_0 = 1$. To find an appropriate value for l_{n+1} having l_n, we prove the following

Lemma 11. There exists a number l' such that every α's segment of length l' contains a good occurrence of every string in B_n.

Proof. Let string x in the alphabet $\{1, 2, \dots, l_n\}$ be $1, 2, \dots, l_n, 1, 2, \dots, l_n$, and a sequence β in the same alphabet to be an infinite concatenation $xxx\dots$. Define the cross product of string of equal lengths similarly to the cross product of infinite sequences. Then u is in B_n iff $u \times x$ has infinitely many occurrences in $\alpha \times \beta$. According to Corollary 5, the sequence $\alpha \times \beta$ is almost periodic, so there exists l' such that every segment of length l' has an occurrence of $u \times x$ for every $u \in B_n$. So, every segment of α of length l' has a good occurrence of every $u \in B_n$. This completes the proof of the Lemma. □

Let l_{n+1} be a number such that $l_n \mid l_{n+1}$ and every α's segment of length l_{n+1} has a good occurrence of every string from B_n.

Let us prove that $\langle l_n, A_n, B_n\rangle$ is a scheme. Condition (C1) is obviously met. The first part of condition (C2) says that every string in B_n consists of two strings from A_n. This is surely true since every good occurrence of the string v_1v_2 has a good occurrence of each of the strings v_1 and v_2. The second part states that every string from A_n is used as part of B_n. If $v_1 \in A_n$, then v_1 has infinitely many occurrences. Consider all strings of length l_n that immediately follow these occurrences. There are finitely many types of these strings, so at least one of them, say, v_2, occurs infinitely many times. Then the string v_1v_2 has infinitely many good occurrences, and thus is in B_n.

To check condition (C3), it is sufficient to prove that if $u \in A_{n+1}$, $u = v_1v_2\dots v_k$ where $|v_i| = l_n$, $k = \frac{l_{n+1}}{l_n}$, then for each $i < k$ $v_iv_{i+1} \in B_n$ and for every string $w \in B_n$ there exists $i < k$ such that $w = v_iv_{i+1}$.

Since $u \in A_{n+1}$, u has infinitely many good occurrences in α. Hence, for all $i < k$ v_iv_{i+1} has infinitely many occurrences in α with a start of the form $cl_{n+1} + (i-1)|v_i|$. But this expression is a multiple of l_n, so v_iv_{i+1} has infinitely many good occurrences in α, so $v_iv_{i+1} \in B_n$ for all $i < k$.

Now suppose $w \in B_n$. The string u has a good occurrence in α (even infinitely many ones). Let one of these be $[j, j+l_{n+1}-1]$. According to the choice of l_{n+1}, the segment $[j, j+l_{n+1}-1]$ has a good occurrence of the string w, so for some i we have $v_iv_{i+1} = w$.

Is remains to check condition (C4). Suppose $u = v_1\dots v_kw_1\dots w_k \in B_n$. Then u has infinitely many good occurrences in α. It follows that v_kw_1 has infinitely many occurrences starting at position which is multiple of l_{n-1} and thus $v_kw_1 \in B_{n-1}$.

Now we prove that α is perfectly generated by the constructed scheme. For every n we let k_n be the multiple of l_n such that every string $u \times x$ that has only finite number of occurrences in $\alpha \times \beta$, does not have any occurrences to the right of k_n.

(b) It is easy to check that the proof in both directions is effective. □

Now we describe the universal method of generating strongly almost periodic sequences. Say that $\langle l_n, A_n\rangle$ is a strong Σ-scheme if for l_n and A_n the property (C1) holds, and for every n every string $u \in A_{n+1}$ is of the form $u = v_1v_2\dots v_k$ where $v_i \in A_n$ and for every $w \in A_n$ there exists $i < k$ such that $w = v_i$. Also, we say that α is generated by a strong scheme if for every i and n $\alpha[il_n, (i+1)l_n - 1] \in A_n$.

The theorem analogous to the Theorem 10 is as follows:

Theorem 12. The sequence α is strongly almost periodic iff it is generated by some strong Σ-scheme.

The proof of this Theorem is analogous to the proof of Theorem 10, although more simple, and is omitted here.

Now we prove that the block product is strongly almost periodic.

Proposition 13. Let u_n be a sequence of $\mathbb{B}$-strings each starting with 0 and containing at least two characters. Then the sequence $\bigotimes\limits_{n=0}^{\infty} u_n$ is generated by some strong $\mathbb{B}$-scheme.

Proof. Let $\alpha = \bigotimes_{n=0}^{\infty} u_n$. Consider two cases.

(a) Starting from some n all the strings u_n do not contain 1. Then α has the form $vvv\ldots$ for some v and thus is periodic. The scheme can be constructed trivially.

(b) For an infinitely many n's the string u_n contains at least one 1. Then α can be represented as $\bigotimes_{n=0}^{\infty} w_n$ where each w_n starts with 0 and contains 1. We prove this by using the associative property of the block product. The product

$$u_0 \otimes u_1 \otimes \ldots \otimes u_n \otimes \ldots$$

can be divided into groups

$$(u_0 \otimes u_1 \otimes \ldots \otimes u_{n_1-1}) \otimes (u_{n_1} \otimes \ldots \otimes u_{n_2-1}) \otimes \ldots$$

so that each group contains and least one term that contains 1. Letting w_i be the block product of the i'th group, we get w_i start with 0 and contain at least one 1.

Now we define the strong $\mathbb{B}$-scheme generating $\alpha = \bigotimes_{n=0}^{\infty} w_n$. Let $x_n = \bigotimes_{i=0}^{n} w_i$, $l_n = |x_n|$, and $A_n = \{x_n, \bar{x}_n\}$. Since for every n the string w_n contains both 0 and 1, $\langle l_n, A_n \rangle$ is a strong $\mathbb{B}$-scheme. It is obvious that α is generated by this scheme.

The proposition is proved. □

4.3 Dynamic systems

Let V be a topological space, $A_1, \ldots, A_k$ be pairwise disjoint open subsets of V, $f\colon V \to V$ be a continuous function, and $x_0 \in V$ be a point such that its orbit $\{f^n(x_0) \mid n \in N\}$ lies inside $\bigcup_{j=0}^{k} A_j$. Define the sequence $\alpha\colon \mathbb{N} \to \{1, \ldots, k\}$ by the condition $f^n(x_0) \in A_{\alpha(n)}$. We will show here two conditions yielding that α is strongly almost periodic and one yielding that α is effectively and strongly almost periodic. (We say that α is effectively and strongly almost periodic if it is computable and given u we can compute n such that either u does not occur in α or every α's segment of length n has an occurrence of u.) We will first formulate the three corresponding theorems and then prove them altogether.

Theorem 14. If V is compact and the orbit of any point of V is dense in V, then α is strongly almost periodic.

Theorem 15. If V is a compact metric space and f is isometric, then α is strongly almost periodic.

It follows from the Theorem 15 that if x/π is irrational, then the sequence $\{$the sign of $\sin nx\}$ is strongly almost periodic: to prove this, one can take a circle for the V and a rotation with the angle x for the f.

Before we formulate the third theorem, fix some definitions. The set $T^s = [0,1)^s$ is called s-dimensional torus. Fix the following metric on T^s. Let the mapping $\phi\colon \mathbb{R}^s \to T^s$ be defined by equality $\phi(x_1, \ldots, x_s) = (\{x_1\}, \ldots, \{x_s\})$ where $\{x\}$ denotes the fractional part of x. Then $\rho(a,b) = \min\{|a' - b'| : \phi(a') = a, \phi(b') = b\}$.

A set $A \subset \mathbb{R}^s$ is called algebraic if it is a solution set of some system of polynomial inequalities (either strict or not) with integer coefficients. A set is called semi-algebraic if it is a union of a finite class of algebraic sets. A set $A \subset T^s$ is called semi-algebraic if there exists a semi-algebraic $B \subset \mathbb{R}^s$ such that $A = B \cap T^s$.

Suppose $v \in \mathbb{R}^s$. The mapping $f_v\colon T^s \to T^s$ defined by the equality $f_v(x) = \phi(x + v)$ is called a shift by the vector v. This mapping is surely isometric.

Theorem 16. Let V be s-dimensional torus, the point x_0 have algebraic coordinates, f a shift by a vector with algebraic coordinates, and A_i open semi-algebraic sets. Then α is effectively and strongly almost periodic.

Proof. (of Theorems 14, 15 and 16) We start with proving Theorem 14. We need to show that if a string $u \in \{1, \ldots, k\}^*$ has an occurrence in α then u is contained in any sufficiently long segment of α. Let u be of length l and have an occurrence in α, say, $u = \alpha[i_0, i_0 + l - 1]$. Denote by B_u the open set

$$\{x \in V \mid x \in A_{u(1)}, f(x) \in A_{u(2)}, \ldots, f^{l-1}(x) \in A_{u(l)}\}.$$

Then $f^{i_0}(x_0) \in B_u$, so B_u is not empty. Since every orbit is dense in V, we have $\forall x \in V\, \exists i \in \mathbb{N}\, f^i(x) \in B_u$. This means $V \subset \bigcup_{i=0}^{\infty} f^{-i}(B_u)$. Since each set $f^{-i}(B_u)$ is open and V is compact, there exists $m \in \mathbb{N}$ such that $V \subset \bigcup_{i=0}^{m} f^{-i}(B_u)$. That is, $\forall x \in V\, \exists i \le m\, f^i(x) \in B_u$. In particular, $\forall n\, \exists i \le m\, f^{n+i}(x_0) \in B_u$, so any α's segment of length $m+l+1$ contains an occurrence of u.

Let us prove Theorem 15 by reduction to Theorem 14. Let V_1 be a closure of the orbit of x_0. Then V_1 is also compact. Denote the metric of V by ρ.

Lemma 17. $f(V_1) \subset V_1$.

Proof. Suppose $x \in V_1$. We prove that $f(x) \in V_1$. Let $\varepsilon > 0$. There exists $k \in \mathbb{N}$ such that $\rho(f^k(x_0), x) < \varepsilon$. Hence $\rho(f^{k+1}(x_0), f(x)) < \varepsilon$ because f is isometric. Since this holds for every $\varepsilon > 0$, $f(x) \in V_1$. □

Lemma 18. For all $x \in V_1$ the orbit of x is dense in V_1.

Proof. Let $x \in V_1$, $y \in V_1$, $\varepsilon > 0$. We need to show that there exists n such that $\rho(f^n(x), y) < \varepsilon$. There exist k and l such that $\rho(f^k(x_0), x) < \varepsilon/3$, $\rho(f^l(x_0), y) < \varepsilon/3$ (since $x, y \in V_1$). We have two cases.

Case 1: $l \ge k$. Take $n = l - k$. We have

$$\rho(f^{l-k}(x), y) \le \rho(f^{l-k}(x), f^l(x_0)) + \rho(f^l(x_0), y) = \rho(x, f^k(x_0)) + \rho(f^l(x_0), y) < \varepsilon/3 + \varepsilon/3 < \varepsilon.$$

Case 2: $l < k$. First we prove that there exists a number $l' \ge k$ such that $\rho(f^{l'}(x_0), f^l(x_0)) < \varepsilon/3$. Then $\rho(f^{l'}(x_0), y) < 2\varepsilon/3$ and we can reason as in case 1.

Since V is compact, for any $\delta > 0$ there exists N such that among any N point there exist two with a distance less than δ. Take N corresponding to $\delta = \frac{\varepsilon}{3k}$. Among the points $f(x_0), f^2(x_0), \ldots, f^N(x_0)$ there are two with a distance less than $\frac{\varepsilon}{3k}$. Let these be $f^{i_0}(x_0)$ and $f^{i_0+r}(x_0)$ (where $r > 0$). Then $\rho(f^{i_0}(x_0), f^{i_0+r}(x_0)) < \frac{\varepsilon}{3k}$, and since f is isometric, for any i we have $\rho(f^i(x_0), f^{i+r}(x_0)) < \frac{\varepsilon}{3k}$. In particular,

$$\begin{gathered}\rho(f^l(x_0), f^{l+r}(x_0)) < \tfrac{\varepsilon}{3k}, \\ \rho(f^{l+r}(x_0), f^{l+2r}(x_0)) < \tfrac{\varepsilon}{3k}, \\ \cdots \\ \rho(f^{l+(k-1)r}(x_0), f^{l+kr}(x_0)) < \tfrac{\varepsilon}{3k},\end{gathered}$$

and hence $\rho(f^l(x_0), f^{l+kr}(x_0)) < \varepsilon/3$. Now we can take $l' = l + kr \ge k$. The proof of the lemma is complete. □

Now we can prove Theorem 15. For the space V_1, the function $f_1 = f|_{V_1}$, the point x_0 and the sets $A_i' = A_i \cap V_1$ all conditions of Theorem 14 hold. Hence α is strongly almost periodic and the Theorem 15 is proved.

Let us switch to proving Theorem 16. Since T^s is a compact metric space and the shift is isometric, the resulting sequence is almost periodic according to Theorem 15. Our goal is effectiveness issues.

Lemma 19. If V is a compact metric space, f is isometric, A_i are open subsets of V, and the following conditions hold (here when we talk of a point in the orbit, it is meant to be represented by its number):

(a) Given a point of the orbit in one of the sets A_i, one can calculate the number i of the set containing this point and a positive rational number ε such that all the point's ε-neighborhood lies in the set A_i.

(b) Given ε, one can effectively find an ε-net[1] in the the orbit of x_0.

(c) Given two points in the x_0's orbit, one can approximate the distance between them.

(d) Given u one can compute if u occurs anywhere in α.

Then, α is effectively and strongly almost periodic.

Proof. Denote $x_n = f^n(x_0)$.

We are given u and we should find such m that every α's segment of length m contains an occurrence of u. Suppose u occurs in α, say, $u = \alpha[p, q]$ (we can find out if it occurs anywhere using (d), and if it

[1]Here under ε-net in the set A we mean a finite set of points $a_i \in A$ such that every point of A is closer than ε to at least one point a_i.

does, find the needed index by trying them in turn). Find the points $x_p, \ldots, x_q$ and for each point x_k find a number ε_k such that all the ε_k-neighborhood of this point is included in the set $A_{\alpha(k)}$ (we can do this using (a)). Let $\varepsilon = \min\{\varepsilon_k\}$ and let $\delta = \varepsilon/4$.

Construct δ-net in the orbit of x_0 using (b). Starting at x_0, start calculating points of the orbit until every point of δ-net is approximated with an error $< \delta$ (here we use (c)). Suppose we needed to calculate l points of the orbit. Then $m = 2l$. Let us prove this.

Suppose we have some segment of α of length m starting at index r. Consider the corresponding points in the orbit, $x_r, \ldots, x_{r+m-1}$. Take the middle point of this segment, x_{r+l}, and find the point y of δ-net that is closer than δ to it. Find the point in the starting segment of α that is closer than δ to y. Suppose it has the number $n < l$. Then the point x_{r+l-n} is closer than 2δ to x_0.

Now perform a similar operation with a point x_p (the starting point of a known occurrence of u). Namely, find a point z in the δ-net that is closer than δ to x_p and find a point in the starting segment of α that is closer than δ to z. Suppose it has the number $s < l$. The point x_s is closer than 2δ to x_p.

Remember that the point x_{r+l-n} is closer than 2δ to x_0. Thus we have that the point $x_{r+l-n+s}$ is closer than 4δ to x_p. Since $4\delta = \varepsilon$, the point $x_{r+l-n+s}$ is closer than ε to x_p, so there is an occurrence of u starting at index $r + l - n + s$.

The lemma is proved. □

Now we need to show that in the situation of Theorem 16 the conditions (a)–(d) of Lemma 19 hold.

One major construct that is used heavily in the following proof is the Tarski Theorem [11]. It states that if we have a first order formula $\psi(x_1, \ldots, x_\sigma)$ in the signature $\{+, \times, <\}$ and representations of algebraic numbers $a_1, \ldots, a_\sigma$, we can find out if $\psi(a_1, \ldots, a_\sigma)$ is true in the ordered field of real numbers. Call a set A representable if there exists a first order formula $\psi(x)$ that is true iff $x \in A$. Surely any semi-algebraic set in the torus is representable.

Also, we need some properties of algebraic numbers. The representation of an algebraic number γ is $\langle Q, a, b\rangle$ where Q is a polynomial with integer coefficients such that $Q(\gamma) = 0$ and $a < b$ is rational numbers such that the interval (a, b) contains γ and does not contain any other root of Q. With this representation, one can effectively add, subtract, multiply and divide algebraic numbers. (It can easily be done using the Tarski theorem.) Also, given a representation of γ one can effectively find a prime polynomial P such that $P(\gamma) = 0$. The proof of this fact is well known. Following is the sketch, for details see, for example, [14].

First, note that if $P = QR$ (where P, Q and R are all polynomials with rational coefficients), then the common denominator of Q's coefficients is less than the common denominator of P's coefficients, and the same holds for R. Then, since the coefficients of a polynomial are symmetric polynomials in its roots, and the set of roots of Q is a subset of the set of roots of P (same for R), the Q's coefficients are bounded in absolute value by some computable functions of the P's coefficients. So, we have only a finite set of possible values for the Q's coefficients. Trying all the possible variants, we understand if there exists a polynomial Q that divides P.

Let us check the conditions.

(a) Given a point with algebraic coordinates (all points in the orbit have algebraic coordinates since both x_0 and the shift vector have algebraic coordinates) we can write a formula $\psi_i(\gamma)$ stating that any point at a distance less than γ is in A_i. Then, enumerating all rational numbers, we can estimate from below the needed neighborhood radius.

(c) All points involved will have algebraic coordinates, so the distance will be algebraic, and thus it can be approximated.

Checking (b) and (d) is harder. We will do this after studying the structure of V_1 (the closure of x_0's orbit) more thoroughly.

Lemma 20. V_1 is a union of a finite number of affine subspaces of equal dimensions.

Proof. Take a point $a \in V_1$. If there exists a neighborhood of a that does not contain any other points of V_1, then the orbit is finite.

Otherwise, there are points in the orbit at deliberately small distances from a. Consider straight lines going through a and these points, and the directions of these lines (in other words, the points where these lines meet a unit sphere centered at a). Since sphere is compact, there is a nonempty set of limit directions. (Such directions w that for every $\varepsilon > 0$ and $\delta > 0$ there exist infinitely many points in the orbit such that they are closer than ε to a and the corresponding directions are closer than δ to w.) Consider the corresponding straight lines. We prove that their affine cull is contained in V_1. Further we will intermix references to V_1 and the corresponding object in $\mathbb{R}^s$ because their connection is trivial and it is generally evident what object is meant.

First, we prove that every limit line is contained in V_1. Take a point $x \in \mathbb{R}^s$ on the line. There exists a point y in the orbit such that $\rho(a,y) < \varepsilon/4$ and the angle between the vectors (a,x) and (a,y) is less than $\frac{\varepsilon}{\mathrm{const}\|x-a\|}$. Also, there exists a point z in the orbit such that $\rho(a,z) < \frac{\varepsilon}{\mathrm{const}}\rho(a,y)$. Then, the angle between (a,x) and (z,y) is still very small (less than $\frac{\varepsilon}{\mathrm{const}\|x-a\|}$).

We need to make sure that z is earlier in the orbit than y. If z is later, we change y as follows. Find a point y' in the orbit later than z such that $\rho(y',y) < \frac{\varepsilon}{\mathrm{const}}\rho(z,y)$, so the angle changes little, and the line (z,y') is still close to (a,x). Let the new y be this y'.

Now we have that the angle between (z,y) and (a,x) is less than $\frac{\varepsilon}{\mathrm{const}\|x-a\|}$, and $\rho(z,y) < \varepsilon/2$. Let us traverse z along the orbit until it becomes y. In the same number of steps y becomes another y_1 such that $y_1 - y = y - z$. So, y_1 lies on the line (z,y). Repeating the operation, we get to the neighborhood of x. The nearest to x point of the sequence y_n is at distance not more than the sum of the distance between x and the line (z,y) (which is less than $\varepsilon/2$ according to our construction) and the distance between two points in the sequence (which is $\rho(z,y) < \varepsilon/2$). So, we have approximated x by the point in the orbit with error not more than ε. This proves that $x \in V_1$.

Up to this point, we know that every limit line is contained in V_1. Our next goal is to prove that their affine cull is contained in V_1. Suppose we proved that a cull of some of the lines is contained in V_1. Take a new limit line that is linearly independent of the considered cull (say, (a,b)) and prove that the new cull is still contained in V_1. Consider a point $x \in \mathbb{R}^s$ in the new cull and project it along (a,b) to the previous cull. Denote the projection x_1. Using the same technique as above, find two points z and y in the orbit that are close to a, and such that the angle between (z,y) and (a,b) is less than $\frac{\varepsilon}{\mathrm{const}\|x-x_1\|}$. Also, we need z to be earlier in the orbit than y. Find a point x_1' in the orbit that is later in the orbit than z and is closer to x_1 than $\varepsilon/2$. Traverse z along the orbit until it becomes x_1'. Then y becomes y'. We have $\rho(y',x_1') < \varepsilon/2$, and the angle between (x_1',y') and (x_1,x) is less than $\frac{\varepsilon}{\mathrm{const}\|x-x_1\|}$. Traversing x_1' to become y' and further, as above, we find a point in the orbit that is closer than ε to x. We just added a new line to the cull. This procedure increases the dimension of the cull, so it can be performed only finitely many times.

Now we prove that all points of the orbit that are not contained in the cull are not closer to the cull than some a positive distance.

Assume for any $\varepsilon > 0$ there exists a point $x(\varepsilon)$ in the orbit that is closer than ε to the cull but is not contained in it. Take $\varepsilon > 0$. Take $x(\varepsilon)$ and a point y in the orbit and in the cull such that y is close to the orthogonal projection of $x(\varepsilon)$. Traverse x and y along the orbit until y becomes some point y' close to a. Then x becomes x' such that (y',x') is almost orthogonal to the cull. Hence (a,x') is almost orthogonal to the cull. As $\varepsilon \to 0$ we have $x' \to a$, and (a,x') tend to be perpendicular to the cull. So, we found a new limit line, contradiction.

Now every point of the orbit is contained in an affine subspace of the same dimension d (since every one of them can be obtained from another by a shift; this also shows that all subspaces are parallel). Consider an orthogonal complement to these subspaces and project them to this complement. Every subspace projects into a point. The distance between any two of these points is more than some positive number. So, there is only a finite number of these affine subspaces. $\square$

Note that if W is one of the affine subspaces such that $W \cap T^s \neq \emptyset$ and $W \cap T^s \subset V_1$, then also $\phi(W) \subset V_1$. This follows from the proof of Lemma 20.

We want to find these affine subspaces given f and x_0. Without loss of generality we can assume that $x_0 = 0$ since we always can shift the origin of the torus to x_0. Let the translation vector v have coordinates $(t_1, \ldots, t_s)$.

Lemma 21. Let $d' = \dim_{\mathbb{Q}}\{t_1, \ldots, t_s, 1\} - 1$. Then the dimension d of the affine subspaces equals d'.

Proof. Recall that d' is the cardinality of the minimal subset of coordinates t_i such that all the coordinates can be rationally expressed in terms of these coordinates and 1.

First, we prove that $d \leq d'$. Without loss of generality, we assume that the first $k - 1 = s - d'$

coordinates $t_1,\dots,t_{k-1}$ can be expressed in terms of the last d': $t_k \dots t_s$. Write these expressions:

$$\begin{aligned} t_1 &= \alpha^1_k t_k + \ldots + \alpha^1_s t_s + \alpha^1_0 \cdot 1, \\ &\dots \\ t_{k-1} &= \alpha^{k-1}_k t_k + \ldots + \alpha^{k-1}_s t_s + \alpha^{k-1}_0 \cdot 1. \end{aligned}$$

Consider these relations for the components of the vector vn. We see that $t'_i = nt_i - m_i \cdot 1$. So the relations are the same except the coefficients α^i_0 differ. If we make the denominator of all fractions α^i_j the same, we will see that the denominator of α^i_0 remains the same when going from f to f^n (this is because m_i are integers). Since all the t'_i are less than 1, the absolute values of coefficients α^i_0 are bounded above. Hence there is only a finite number of possible values for α^i_0. So, for any n the vector vn that is equal to $f^n(x_0)$ (since $x_0 = 0$) lies in one of the finite number of affine subspaces of dimension d':

$$\begin{aligned} T_1 &= \alpha^1_k T_k + \ldots + \alpha^1_s T_s + \beta^1_j \\ &\dots \\ T_{k-1} &= \alpha^{k-1}_k T_k + \ldots + \alpha^{k-1}_s T_s + \beta^{k-1}_j \end{aligned}$$

(here T_i are coordinates and β^i_j is the j'th possible value for α^i_0). Hence $d \le d'$.

Now we prove that $d \ge d'$. Project the whole picture onto the last d' coordinates $k,\dots,s$. If $d < d'$ then each affine subspace of V_1 projects into subspace of dimension not more than d, so they all cannot cover the whole coordinate subspace. Let us prove that the projection of V_1 covers all the subspace generated by the coordinates $k,\dots,s$.

More precisely, we prove the following: if we project the whole picture onto a coordinate subspace of dimension $l \le d'$, the image will cover all the mentioned subspace. We do this by induction on l. The induction base is $l = 0$. This case is obvious. Assume we proved the statement with some value of l. Let us prove it with $l+1$. Project the picture onto last l coordinates. According to the induction hypothesis, the image has the dimension l. So, the projection onto the last $l+1$ coordinates has a dimension of either $l+1$ or l. We need to prove that it is $l+1$. Assume, for the contrary, that the dimension is l, that is, the projection of V_1 is a union of parallel affine subspaces of dimension l. They are not parallel to any coordinate axis (because if they were, we could project the picture along this axis, and the spaces would project into spaces of dimension $l-1$, which cannot be true due to the induction hypothesis). The subspaces intersect s'th coordinate axis by a point. The distances between adjacent points are the same. Since the coordinate axis can be regarded as a circle (because we are in the torus!), this distance is rational. Write the equation of j'th subspace

$$t_s = \alpha'_{s-l} t_{s-l} + \ldots + \alpha'_{s-1} t_{s-1} + \beta'_j.$$

Since for different j the difference between β'_j is rational, and the point 0 is contained in one of them, then all β'_j are rational.

Consider the subspace containing 0 and its intersection with a two-dimensional coordinate subspace of coordinates s and q (where $q \ge s-l$). Its equation is $t_s = \alpha'_q t_q$. Consider a vector in this subspace (but outside the torus) with q-coordinate of 1. Denote its s-coordinate by x_s. We have

$$x_s = \alpha'_q \cdot 1.$$

The equivalent vector in the torus has q-coordinate of 0, and s-coordinate of $x_s - m$ for some integer m. It is contained in some affine subspace number j, so

$$x_s - m = \alpha'_q \cdot 0 + \beta'_j.$$

Since β'_j is rational, then the number

$$\alpha'_q = \beta'_j + m$$

is rational too. So, all the coefficients α'_q are rational. This contradicts the fact that $\{t_k,\dots,t_s,1\}$ are linearly independent over $\mathbb{Q}$. $\square$

Now we are ready to prove that the conditions (b) and (d) of Lemma 19 hold in our case.

First, find a primitive element γ in the field $\mathbb{Q}[t_1,\dots,t_s,(x_0)_1,\dots,(x_0)_s]$, represent all coordinates of the vectors v and x_0 as polynomials in γ and find $d = d'$ and the coefficients of all equations of affine subspaces—except for the coefficients β_j^i (remember the beginning of the proof of Lemma 21). We can find all possible values for β_j^i, but we still need to know which give us the needed subspaces of V_1. To find these, we compute $x_0, x_1, \dots, x_N$ (note that we write x_n for $f^n(x_0)$). The number N is chosen such that these points constitute a ε-net (for some sufficiently small ε) in every subspace that has at least one point of $x_0, \dots, x_{N+1}$. Then we can say that we have all the subspaces. Suppose we then jump (at n'th step) from a known subspace to a not yet known. There was a point x_m of the ε-net near to x_n. Then there is a point x_{m+1} near to x_{n+1}. But x_{n+1} is in the new subspace, and $\rho(x_{m+1}, x_{n+1}) = \rho(x_m, x_n) < \varepsilon$, so x_{m+1} is also in the new subspace (remember that subspaces are separated by a positive distance), so really this subspace is not new, but old.

Hence we can find the closure of the orbit and thus build a ε-net in it. So, the condition (b) is met. Knowing V_1, we can also meet the condition (d). Suppose we have a string u and want to know if it occurs anywhere in the sequence α. We construct the set

$$B_u = \{y \mid y \in T^s, \phi(y) \in A_{u(1)}, \dots, \phi(y + (|u| - 1)v) \in A_{u(|u|)}\}$$

This set is representable since A_i is semi-algebraic sets and v has algebraic coordinates. We can, given u, v and A_i, find a formula $\psi(x)$ that is true iff $x \in B_u$. Then, we can construct a formula stating that there is a point y in the closure of the orbit such that $y \in B_u$. Then, we use the Tarski theorem to find out if there exists such point. So, the condition (d) is also met, and this, finally, proves the Theorem 16. □

5 Interesting examples

Theorem 22. For any $m \in \mathbb{N}$ there exists a set A of $m + 1$ effectively almost periodic $\mathbb{B}$-sequences such that the cross product of any m sequences from A is effectively almost periodic, and the cross product of all $m + 1$ sequences is not almost periodic.

Theorem 23. For any $m \in \mathbb{N}$ there exists a set A of $m + 1$ effectively almost periodic $\mathbb{B}$-sequences such that the cross product of any m sequences from A is effectively almost periodic, and the cross product of all $m + 1$ sequences almost periodic but not effectively almost periodic.

A homomorphism $h\colon \Sigma^* \to \Delta^*$ is called a collapse if for any character $\sigma \in \Sigma$ $|h(\sigma)| = 1$ and $|\Delta| < |\Sigma|$.

Theorem 24. For any $m \in \mathbb{N}$ there exists a computable sequence $\alpha\colon \mathbb{N} \to \{1, \dots, m\}$ such that for any collapse h the sequence $h(\alpha)$ is effectively almost periodic. Such sequence can be constructed to conform to one of the two conditions:

(a) α is not almost periodic,

(b) α is almost periodic, but not effectively almost periodic.

Proof. (of Theorems 22, 23 and 24) We say that $\langle l_n, A_n, B_n\rangle$ is pseudoscheme if for any collapse h the tuple $\langle l_n, h(A_n), h(B_n)\rangle$ is a scheme. We start by proving Theorem 24(a). To do this, we construct a pseudoscheme $\langle l_n, A_n, B_n\rangle$ and a non-almost periodic sequence α such that for any collapse h $h(\alpha)$ is effectively generated by $\langle l_n, h(A_n), h(B_n)\rangle$.

Let Σ_m be the alphabet $\{1, \dots, m\}$. We will identify permutations over Σ_m with strings of length m in the alphabet Σ_m without equal characters.

Define a sequence l_n and auxiliary sets $R_n^u \subset \Sigma_m^{l_n}$ (where $u \in \mathbb{B}^{n+1}$). The sets R_n^u for different $u \in \mathbb{B}^{n+1}$ should be pairwise disjoint and have equal cardinalities.

We let l_0 be m, R_0^0 be the set of even permutations over Σ_m, and R_0^1 be the set of odd permutations over Σ_m.

Suppose l_n and the sets R_n^u are already defined so that the sets R_n^u are pairwise disjoint and have equal cardinalities. Denote $O_n^v = R_n^{v0} \cup R_n^{v1}$ for all $v \in \mathbb{B}^n$. We say that the string u is a complete concatenation of strings for a finite set M if $u = v_1 v_2 \dots v_k$ is a concatenation of strings from M such that for every two strings $w_1, w_2 \in M$ there exists an index $i < k$ such that $w_1 = v_i$ and $w_2 = v_{i+1}$. Let k_{n+1} be a minimal k such that there exists a complete concatenation of strings from O_n^v (since O_n^v have equal cardinalities, k_n does not depend on v). Let $l_{n+1} = l_n(k_{n+1} + 2)$.

For $u \in \mathbb{B}^{n+2}$ we define R^u_{n+1} as follows. Let ε, δ be the last two characters of u so that $u = u'\varepsilon\delta$. Let

$$R^u_{n+1} = \{v_1 \dots v_{k_{n+1}} w_1 w_2 \mid v_1 \dots v_{k_{n+1}} \text{ is a complete concatenation from } O^{u'}_n, w_1 \in R^{u'\varepsilon}_n, w_2 \in R^{u'\delta}_n\}$$

It is obvious that R^u_{n+1} are pairwise disjoint and have equal cardinalities. We will name O^v_n zones of rank n and R^u_n regions of rank n. So, $R^{v\varepsilon}_n$ is a region of zone O^v_n when $\varepsilon \in \mathbb{B}$. We thus have 2^n pairwise disjoint zones of rank n, each being a disjoint union of two regions of rank n.

Let $\tau = v_0, v_1, \dots$ be a sequence of $\mathbb{B}$-strings such that $|u_n| = n$. Let $A^\tau_n = O^{v_n}_n$, and let B^τ_n be $A^\tau_n A^\tau_n$, a set of pairwise concatenations of strings from A^τ_n. We prove that $\langle l_n, A^\tau_n, B^\tau_n\rangle$ is a pseudoscheme.

Lemma 25. For any collapse h, for any n and any string u_1, u_2 of length $n+1$ there exists a bijection $\phi\colon R^{u_1}_n \to R^{u_2}_n$ such that $\forall x \in R^{u_1}_n\ h(x) = h(\phi(x))$ (in particular, $h(R^{u_1}_n) = h(R^{u_2}_n)$).

Proof. We use induction over n.

Let $n = 0$. If $u_1 = u_2$, let ϕ be an identity function. If $u_1 = 0$, $u_2 = 1$, we take $i, j \in \Sigma_m$ such that $h(i) = h(j)$ (such i and j do exist because h is a collapse). Define ϕ by the equalities $\phi(i) = j$, $\phi(j) = i$, and $\phi(k) = k$ for $k \neq i, j$.

Suppose the statement for n is already proved. Then for any $u_1, u_2 \in \mathbb{B}^n$ there exists a bijection $\phi\colon O^{u_1}_n \to O^{u_2}_n$ that preserves h. We construct a bijection for any two regions of rank $n+1$. Let $u_1\varepsilon_1\delta_1$ and $u_2\varepsilon_2\delta_2$ be any two strings of length $n+2$, where $|u_i| = n$, $\varepsilon_i, \delta_i \in \mathbb{B}$. Then every string in $R^{u_1\varepsilon_1\delta_1}_{n+1}$ can be represented as $x = v_1 \dots v_{k_{n+1}} w_1 w_2$ where $v_i \in O^{u_1}_n$, $w_1 \in R^{u_1\varepsilon_1}_n$, $w_2 \in R^{u_1\delta_1}_n$. By the induction hypothesis, there exist bijections $\phi_1 : O^{u_1}_n \to O^{u_2}_n$, $\phi_2 : R^{u_1\varepsilon_1}_n \to R^{u_2\varepsilon_2}_n$, and $\phi_3 : R^{u_1\delta_1}_n \to R^{u_2\delta_2}_n$, that preserve h. Let

$$\phi(x) = \phi_1(v_1)\phi_1(v_2)\dots\phi_1(v_{k_{n+1}})\phi_2(w_1)\phi_3(w_2).$$

Then $\phi_1(v_1)\dots\phi_1(v_{k_{n+1}})$ is a complete concatenation of strings in $O^{u_2}_n$, thus $\phi(x) \in R^{u_2\varepsilon_2\delta_2}_{n+1}$. Obviously, ϕ is a bijection from $R^{u_1\varepsilon_1\delta_1}_{n+1}$ to $R^{u_2\varepsilon_2\delta_2}_{n+1}$.

Since ϕ_1, ϕ_2 and ϕ_3 preserve h, so does ϕ. $\square$

It follows from this Lemma that the images of all zones under any collapse h coincide, i.e. $h(O^{u_1}_n) = h(O^{u_2}_n)$. It is now obvious that $\langle l_n, h(A^\tau_n), h(B^\tau_n)\rangle$ is a scheme for any τ and h.

Now we construct a sequence of $\mathbb{B}$-strings $\tau = v_0, v_1, \dots$ and non-almost periodic sequence α such that for any collapse h the scheme $\langle l_n, h(A^\tau_n), h(B^\tau_n)\rangle$ effectively generates $h(\alpha)$. Let

$$v_n = \begin{cases} 0^n, & \text{if } n \text{ is even,} \\ 10^{n-1}, & \text{if } n \text{ is odd.} \end{cases}$$

For every $n \in \mathbb{N}$ choose a string x_n from $A^\tau_n = O^{v_n}_n$ and let

$$\alpha = x_0 x_1 \dots x_n \dots.$$

Denote the starting index of x_n by s_n (so, $x_n = \alpha[s_n, s_n + l_n - 1]$).

Let us prove that α is not almost periodic. Suppose it is almost periodic.

It is easy to check that for every $\varepsilon \in \mathbb{B}$ every string in $O^{u\varepsilon}_{n+1}$ is a concatenation of strings from O^u_n. So, for every n the string x_n can be regarded as a concatenation of strings from either $O^{00\dots0}_{n'}$ or $O^{10\dots0}_{n'}$ for any $n' < n$ (the choice depends on the evenness of n).

Every string in $O^{10\dots0}_n$ is a concatenation of strings from O^1_1 (let us call them blocks). For $n \geq 2$ every string from $O^{10\dots0}_n$ contains every string from O^1_1 among its blocks. So, every string from O^1_1 has infinitely many occurrences in α.

Consider one of these occurrences, say, $[i, j]$. Call this occurrence nice if $i \equiv s_1 \pmod{l_1}$. We can see that every occurrence of a string from O^1_1 as a block in some x_n is always nice. So, every string from O^1_1 has infinitely many nice occurrences. Fix one such string y. It has the form

$$y = v_1 \dots v_{k_1} w_1 w_2$$

where $v_j \in O^\Lambda_0$, $w_1 \in R^1_0$, $w_2 \in R^0_0 \cup R^1_0 = O^\Lambda_0$. Using an argument analogous to that in the proof of Lemma 11, we can show that y has a nice occurrence on every sufficiently long segment of α. So, the string

y has a nice occurrence within every x_n for a sufficiently large n, that is, there is a block in x_n equal to y. Let us show that y cannot be a block of x_n for even n. Since for even n the string x_n is in $O_n^{00\ldots0}$, all the blocks are from O_1^0, that is, they have the form

$$t_1^1 \ldots t_{k_1}^1 r_1^1 r_2^1$$

where $t_j \in O_0^\Lambda$, $r_1 \in R_0^0$, $r_2 \in R_0^0 \cup R_0^1 = O_0^\Lambda$. Hence we have $w_1 = r_1$ which obviously is a contradiction since w_1 is an odd permutation and r_1 is an even one.

Part (a) of Theorem 24 is proved.

Now turn to the part (b). Fix some enumerable, but undecidable set $E \subset \mathbb{N}$. Define a sequence of $\mathbb{B}$-strings v_n as follows. Let $|v_n| = n$ and let $v_n(i) = 1$ if the number i is generated in less than n steps of enumerating E. Then v_n is a computable sequence having the following property: for every i there exists L such that for all $n \geq L$ $v_n(i) = E(i)$, but L cannot be computed given i. Let $A_n = O_n^{v_n}$, and $B_n = A_n A_n$. Then, as it was shown above, $\langle l_n, A_n, B_n\rangle$ is a pseudoscheme. Let (as above)

$$\alpha = x_0 x_1 \ldots x_n \ldots$$

where x_n is lexicographically first string in A_n. It is clear that α is computable. For any collapse h $h(\alpha)$ is effectively generated by $\langle l_n, h(A_n), h(B_n)\rangle$, so $h(\alpha)$ is effectively almost periodic.

Let us show that α is almost periodic. Let e_n be n'th prefix of a characteristic sequence of E, that is, $|e_n| = n$, and $e_n(i) = E(i)$. Take $C_n = O_n^{e_n}$ and $D_n = C_n C_n$. Then $\langle l_n, C_n, D_n\rangle$ is a scheme because $e_{n+1} = e_n E(n)$ and every string in $O_{n+1}^{e_n E(n)}$ is a complete concatenation of strings from $O_n^{e_n}$. Let us prove that α is generated by the scheme $\langle l_n, C_n, D_n\rangle$. Take $n \in \mathbb{N}$. We need to find $m \in \mathbb{N}$ such that for all $j \in \mathbb{N}$ $\alpha[m+jl_n, m+(j+2)l_n-1] \in D_n$. There exists $M \geq n$ such that for all $i \geq M$ v_i starts with e_n. Hence x_i is a concatenation of strings from $O_n^{e_n} = C_n$. It follows that for all $j \in \mathbb{N}$ we have $\alpha[m+jl_n, m+(j+1)l_n-1] \in C_n$, and $\alpha[m+jl_n, m+(j+2)l_n-1] \in D_n$ for some m.

Let us prove that α is not effectively almost periodic. Assume α is effectively almost periodic. We will obtain that E is decidable then. This will easily follow from this property of α: e_n is a unique string such that every string from $O_n^{e_n}$ has infinitely many nice occurrences in α. (Here the word "nice" means that the start position of the occurrence is equal to s_n modulo l_n.) Let us prove this property.

For a sufficiently large i the string v_i starts with e_n, so x_i contains every string from $O_n^{e_n}$, and so α has infinitely many nice occurrences of these strings. If some $w \neq e_n$, denote by j the number of the first character where they differ. Then for a sufficiently large i the string v_i starts with $e_n[0,j]$, and x_i is a concatenation of strings from $O_{j+1}^{e_n[0,j]}$. Using the same technique we used for proving the part (a), one can prove that a string from $O_{j+1}^{w[0,j]}$ cannot be a nice substring of a concatenation of strings from $O_{j+1}^{e_n[0,j]}$. Hence, α contains only a finite number of nice occurrences of strings from O_n^w.

The Theorems 22 and 23 follow from the Theorem 24.

Let us construct a sequence α in the alphabet $\mathbb{B}^{m+1}$ that is not almost periodic, but becomes effectively almost periodic under every collapse. Let α_i be i'th projection in the cross product $\mathbb{B} \times \mathbb{B} \times \ldots \times \mathbb{B}$, having $\alpha = \alpha_1 \times \ldots \times \alpha_{m+1}$. Then the cross product of every m sequences from the set $\{\alpha_1, \ldots, \alpha_{m+1}\}$ results from a collapse of α, and is effectively almost periodic.

Theorem 23 is proved in a similar way. $\square$

6 Almost periodic sequences and Kolmogorov complexity

In this section we study the connection between almost periodicity and Kolmogorov complexity. For the definition see [15]. Here we consider simple complexity $K(x)$.

Let α be an almost periodic sequence and α_n its prefix of length n. We shall study $K(\alpha_n)$ as a function of n.

Consider the following simple example: divide a circle into k arcs with k points (with computable coordinates). Take a real number ϕ such that $\frac{\phi}{2\pi}$ is irrational. Define $\alpha(i)$ as the number of arc containing the point $i\phi$. (Note that $i\phi$ can be one of the delimiting points. However, this can happen only a finite number of times. So, we can think that this does not happen at all.) Then, the constructed sequence α is almost periodic according to Theorem 15.

Theorem 26. For the constructed sequence α,

$$K(\alpha_n) \le \mathcal{O}(\log n)$$

Proof. Denote the division points by $x_1, \dots, x_k$. For every n mark every point on the circle with the number of arc it will go to after being multiplied by n. We will have nk arcs corresponding to the k arcs of initial picture. Call them n-arcs. To tell what arc will contain $n\phi$ it is sufficient to know what n-arc contains ϕ.

Now to describe the n'th prefix of α we can use the numbers of m-arcs containing ϕ for all $m \le n$. To know all these numbers mark the boundaries of all m-arcs for all $m \le n$. There are $\frac{n(n+1)}{2}k$ boundaries. They divide the circle in $\frac{n(n+1)}{2}k$ pieces. We need to know the piece containing ϕ. To write its number, we need $\mathcal{O}(\log(\frac{n(n+1)}{2}k))$ bits.

The program that prints α_n incorporates this number and the number n. Let us describe how it works. It needs to calculate the picture of the boundaries. Since the coordinates $x_1, \dots, x_k$ are computable, we can only estimate the boundaries, and not calculate them precisely. So, for any two boundaries the program estimates them (with higher and higher precision) until it understands that one of them is larger than another. The only problem is that some boundaries can be equal — in this case the algorithm will never stop. So, we need to include the description of these cases in the algorithm. The collision between x_{i_1} and x_{i_2} happens if for some integers a_1, a_2 and a_3 we have

$$a_1 x_{i_1} = a_2 x_{i_2} + a_3 \pi.$$

For any i_1 and i_2 the triples (a_1, a_2, a_3) form a subgroup in $\mathbb{Z}^3$. This subgroup is generated by at most three vectors (for proof see [14]). So, the program will also incorporate these vectors for all pairs (i_1, i_2). When it needs to know if two particular boundaries coincide, it uses the corresponding vectors and gets the answer since the first-order theory of $\langle \mathbb{Z}, + \rangle$ is decidable. The length of the descriptions for the vectors is constant in n.

The length of the program is $\log n + \mathcal{O}(\log(\frac{n(n+1)}{2}k)) + \mathcal{O}(1)$ (the last term is the length of the invariant section). Since $\log(\frac{n(n+1)}{2}k) \le 2\log n + \log k$, we have

$$K(\alpha_n) \le \mathcal{O}(\log n).$$

The proof is complete. $\square$

For simplicity, we will stick to the alphabet $\mathbb{B}$. It is evident that $K(\alpha_n) \le n + \mathcal{O}(1)$ (we can incorporate α_n itself in the program). The following theorem shows that this bound cannot be reached for an almost periodic sequence.

Theorem 27. For any almost periodic sequence α there exists a positive ε such that

$$K(\alpha_n) < (1-\varepsilon)n + \mathcal{O}(1)$$

Proof. First, prove that there exists a string of type I (occurring in α only finitely many times). Either the string 1 or the string 0 belongs to type II. We assume, without loss of generality, that this is the string 0. There exists a number l such that every substring of α of length l contains at least one zero. Thus, a string consisting of $l+1$ 1's occurs only finitely many times. Let u be a string of minimal length that occurs in α only finitely many times. Choose an index q such that there is no occurrence of u to the right of q. From now on, we will consider only the portion of α to the right of q.

If $|u| = 1$ (which implies that α consists entirely of ones or zeroes), then $K(\alpha_n) \le \mathcal{O}(\log n)$, because α_n is effectively determined only by n, and we can incorporate n in the program using $\mathcal{O}(\log n)$ bits.

Let u' be a string resulting when we omit the last character in u. Assume w.l.o.g. that we omitted 0, so $u = u'0$. We know that every occurrence of u' is followed by 1. The string $u'1$ occurs infinitely many times in α (because if it had only finitely many occurrences, u' would have had only finitely many occurrences, which contradicts the assumption that u is the shortest string occurring only finitely many times). Hence there exists m such that every α's substring of length m contains at least one instance of $u'1$.

Let us show a "compression" algorithm that will encode α_n using $(1-\varepsilon)n + \mathcal{O}(1)$ bits. Divide α_n into blocks in the following way: first block has length q and is written directly; following blocks have lengths m

and are encoded; the last block of length m' less than m is also written directly. The encoding procedure finds the first occurrence of $u'1$ in the block and writes the block replacing this occurrence of $u'1$ with u'.

Now we need to show that this encoding does not lose information (i.e. the original string can be effectively reconstructed) and that we can build a program that outputs α_n and has length less than $(1-\varepsilon)n + \mathcal{O}(1)$.

The decoding procedure is obvious. The first block of length q is just left as it is. For every encoded block (it has length $m-1$ because exactly one occurrence of $u'1$ was replaced with u') we find the first occurrence of u' and insert a 1 after it. Finally, the last incomplete block is also left as it is.

Now let us calculate the length of the program to output α_n. It will contain the first and the last blocks of the encoded string, the string u, the number m, and the encoded blocks. The length of the program excluding the encoded blocks is bounded from above by a constant. In the remaining part for every m characters in α we write only $m-1$ bits. So, for $n-q-m'$ characters we will need $(n-q-m')\frac{m-1}{m}$ bits. Thus

$$K(\alpha_n) \leq (n-q-m')\frac{m-1}{m} + \mathcal{O}(1) \leq n\left(1 - \frac{1}{m}\right) + \mathcal{O}(1).$$

This proves the theorem. □

We will show that for every $\varepsilon > 0$ there exists a strongly almost periodic sequence α such that $K(\alpha_n) > n(1-\varepsilon)$. This result is proved in the remaining part of this section, namely,

Theorem 28. For any $\varepsilon > 0$ there exists a strongly almost periodic sequence α such that

$$K(\alpha_n) \geq (1-\varepsilon)n + \mathcal{O}(1)$$

for all n.

Actually, it is sufficient to prove this with $\mathcal{O}(\log n)$ additional term. Indeed, if we have done this, then by decreasing ε we get also $\mathcal{O}(1)$, since $\delta n > C \log n$ for large n.

6.1 The construction

Let us build a scheme $\langle l_n, A_n \rangle$ that will generate our sequence.

Define A_0 to be the set of all strings of length l_0. Let

$$A_n = \{v_1 \dots v_{k_n} \mid v_i \in A_{n-1}, \quad \forall a \in A_{n-1}\, \exists i\colon a = v_i\},$$

where $k_n = \frac{l_n}{l_{n-1}}$. The values for k_n (and for l_n, respectively) as well as for l_0, will be chosen later.

First, we prove the following Lemma:

Lemma 29. Let A be an alphabet. Denote by B the set of all strings of length k that contain all characters in A. Then for any $\varepsilon > 0$, and sufficiently large k the following holds:

$$|B| \geq (1-\varepsilon)|A|^k.$$

Proof. Let us take a random k-character string in the alphabet A and calculate the probability of it containing not all characters of A. It is composed of $|A|-1$ different characters, and

$$\Pr(\text{the string does not contain } i\text{'th character}) =$$
$$\frac{(|A|-1)^k}{|A|^k} = \left(1 - \frac{1}{|A|}\right)^{|A|\frac{k}{|A|}} \leq 2e^{-\frac{k}{|A|}}.$$

Making k very large, we easily obtain

$$\Pr(\text{the string does not contain } i\text{'th character}) \leq \frac{\varepsilon}{|A|},$$

and

$$\Pr(\text{the string contains not all characters}) \leq \varepsilon.$$

Hence, at least a $(1-\varepsilon)$ fraction of strings in A^k are in B, so $|B| \geq (1-\varepsilon)|A|^k$. □

The scheme is built in a way such that

$$|A_n| \geq (1-\varepsilon_n)|A_{n-1}|^{k_n}.$$

We can achieve this due to the last Lemma for any values for ε_n. We will determine these values later.

The sequence α that is generated by this scheme is constructed in the following way. Consider a set F of all sequences α such that

$$\alpha[il_n, (i+1)l_n] \in A_n \tag{1}$$

for all i, n.

Consider also a probabilistic distribution p on the space of all sequences in the alphabet A that is uniformly distributed over the set F. The sequence that has complex prefixes is chosen randomly with respect to p. According to the Levin-Schnorr theorem (see [13]), if α is random with respect to p, then

$$KM(\alpha[0,n]) \geq -\log p(\Gamma_{\alpha[0,n]}) + \mathcal{O}(1),$$

where $\Gamma_{\alpha[0,n]}$ is a cone at $\alpha[0,n]$, i.e. a set of all sequences β such that $\beta[0,n] = \alpha[0,n]$, and KM is a Kolmogorov monotone complexity (see [15]). Since $KM(x) \leq K(x) + \mathcal{O}(\log|x|)$, this gives us the desired result if we prove that $-\log p(\Gamma_{\alpha[0,n]}) \geq (1-\varepsilon)n$.

To prove this, we consider a sequence of distributions $p_0, p_1, \ldots$. Let p_0 be a uniform distribution. Let p_j be a distribution that is uniform over the set of sequences satisfying the condition (1) for all i and all $n \leq j$. Obviously $p_j \to p$ as $j \to \infty$. First, let us consider the transition from p_{j-1} to p_j.

We need to compute the change in probability of $\Gamma_{\alpha[0,n]}$. To do so, we first take $n = l_j$ and look at $\Gamma_{\alpha[0,l_j]}$ under p_{j-1}. Consider the sets Γ_x for $|x| = n$. Some of them (those that correspond to x's which do not conform to the condition in (1)) have zero probability, while others' probabilities are equal. Under p_j some of the sets Γ_x lose their probability due to the fact that their x's do not conform to the new condition, and the others' probabilities increase (but they are still equal among the sets with non-zero probabilities). Namely, there were $|A_{j-1}|^{k_j}$ strings that conformed to the conditions of step $j-1$, and only $|A_j|$ strings that conform to the conditions of step j. Since

$$|A_j| \geq (1-\varepsilon_j)|A_{j-1}|^{k_j},$$

the amount of increase in probability is not more than $\frac{1}{1-\varepsilon_j}$.

If $l_j \mid n$, then obviously the probability increases that amount for each block of length l_j, so the total amount is $\left(\frac{1}{1-\varepsilon_j}\right)^{\frac{n}{l_j}}$.

Now consider the case when $l_j \nmid n$. Denote by t the least multiple of l_j larger than n. For any x the set Γ_x contains $\Gamma_{x'}$ for each x' that continues x and has the length of t. Under p_j some of these sets lose their probability and some increase, but not more than $\left(\frac{1}{1-\varepsilon_j}\right)^{\frac{t}{l_j}}$ times. So, the amount of increase in probability of Γ_x is not more than

$$\left(\frac{1}{1-\varepsilon_j}\right)^{\frac{t}{l_j}} = \left(\frac{1}{1-\varepsilon_j}\right)^{\left\lceil\frac{n}{l_j}\right\rceil}.$$

Combining the results, and taking the product over $j = 0, \ldots$, we obtain

$$p_\infty(\Gamma_{\alpha[0,n]}) \leq p_0(\Gamma_{\alpha[0,n]}) \left(\frac{1}{1-\varepsilon_1}\right)^{\left\lceil\frac{n}{l_1}\right\rceil} \cdots \left(\frac{1}{1-\varepsilon_j}\right)^{\left\lceil\frac{n}{l_j}\right\rceil} \cdots.$$

Since $\left\lceil\frac{n}{l_j}\right\rceil \leq \frac{n}{l_j} + 1$, the bound can be rewritten as

$$\underbrace{\left(\frac{1}{1-\varepsilon_1}\right)\cdots\left(\frac{1}{1-\varepsilon_j}\right)\cdots}_{C} \times \underbrace{\left(\left(\frac{1}{1-\varepsilon_1}\right)^{\frac{1}{l_1}}\cdots\left(\frac{1}{1-\varepsilon_i}\right)^{\frac{1}{l_j}}\cdots\right)^{n}}_{D^n},$$

where C and D are constant factors. Here, C can be made deliberately close to 1 by choosing values for ε_j, and $D \leq C$ since $\frac{1}{1-\varepsilon_j} > 1$ and $\frac{1}{l_j} < 1$. So,

$$p_\infty(\Gamma_{\alpha[0,n]}) \leq p_0(\Gamma_{\alpha[0,n]})C^{n+1} = 2^{-n}C^{n+1} = 2^{-n+(n+1)\log C},$$

and thus

$$-\log p_\infty(\Gamma_{\alpha[0,n]}) \geq n - (n+1)\log C \geq n(1 - 2\log C).$$

Since C can be made deliberately close to 1, $\log C$ can be made deliberately small, and we finally obtain

$$KM(\alpha[0,n]) \geq -\log p_\infty(\Gamma_{\alpha[0,n]}) + \mathcal{O}(1) \geq n(1-\varepsilon)$$

for any $\varepsilon > 0$, which is exactly what we wanted.

7 Acknowledgements

The authors would like to thank Nikolai Vereshchagin for writing initial text of this paper and Alexander Shen for help and suggestions.

References

[1] Yu. L. Ershov. Decidability problems and the constructive models. // Moscow, Nauka, 1980.

[2] S.Kakutani. Ergodic theory of shift transformations // Proc. V. Berkely Simp. Prob. Stat., vol. II, part 2, 1967, p.407-414.

[3] M. Morse and G. A. Hedlund, Symbolic dynamics I, Amer. J. Math 60(1938), pp. 815–866.

[4] M. Morse and G. A. Hedlund, Symbolic dynamics II — Sturmian trajectories, Amer. J. Math 62(1940), pp. 1–42.

[5] M. Sipser. Introduction to the theory of computation. PWS Publishing Company, Boston, 1997. Part 1, pp. 31–123.

[6] Andreas Weber. On The Valuedness of Finite Transducers. Acta Informatica, 27(8), pp. 749–780 (1989).

[7] M.Keane. Generalized Morse sequences // Z. Wahrseheinlichkeitstheorie verw. Geb., 1968, Bd 22, S. 335-353.

[8] R.Loos. Computing in Algebraic Extensions // Compting, 1982, Suppl.4, p.173-187

[9] M.Morse. Recurrent geodesies on a surface of negative curvature // Trans.Amer.Math.Soc. 1921, v.22, p. 84-100.

[10] Semenov A.L. Logic theories of unary functions over natural numbers. Izv. AN SSSR. Ser. matem., 1983, vol. 47, n. 3, pp. 623–658 (Russian).

[11] A.Tarski. A decision method for elementary algebra and geometry. Berkly and Los-Angeles, 1951.

[12] K.Jacobs. Maschinenerzeugte 0-1-Folgen // Selecta Mathematica II.
Springer Verlag: Berlin, Heidelberg, New York. 1970.

[13] Uspensky V.A., Semenov A.L., Shen' A.Kh. Can an individual sequence of zeroes and ones be random? Russian Math. Surveys, 45(1): 121–189, 1990.

[14] B.L. Van der Waerden. Algebra. Springer-Verlag, 1991.

[15] V. A. Uspensky, A. Kh. Shen. Relations between varieties of Kolmogorov complexities. Math. Systems Theory, 29(1996), pp. 271–292.