



Soviet-era science, translated into English

ON SOME PROPERTIES OF CODE SYSTEMS

1961

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-196101.33422>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

CYBERNETICS AND CONTROL THEORY

V. I. LEVENSHTEIN

ON SOME PROPERTIES OF CODE SYSTEMS

(Presented by Academician P. S. Novikov on 24 V 1961)

In the present paper effective criteria are studied for recognizing certain properties⁽¹⁻⁶⁾ of code systems. Algorithms simpler than the known ones^(5,6) are constructed for determining whether a code system has the property of unique decipherability or the property of bounded delay; an algorithm is also found for determining whether a code system has the synchronization property.

1. Let $B = \{b_1, \dots, b_r\}$ be some alphabet. An arbitrary (ordered) set $V = \{v_1, \dots, v_m\}$, consisting of nonempty words in the alphabet B , will be called a **dictionary** over B . The length of a word β in the alphabet B will be denoted by $\lambda(\beta)$. A **code system** $\{A, U; B, V\}$ will mean the totality of a dictionary $U = \{u_1, \dots, u_m\}$ over some alphabet $A = \{a_1, \dots, a_q\}$ and a dictionary $V = \{v_1, \dots, v_m\}$ over some alphabet $B = \{b_1, \dots, b_r\}$. The words $u_i \in U$ will be called **elementary messages**, and the words $v_i \in V$ **elementary codes**. Words in the alphabet A representable in the form of products $u_{i_1} \dots u_{i_k}$ will be called **messages**, and words in the alphabet B representable in the form of products $v_{j_1} \dots v_{j_l}$ will be called **codes**. The one-to-one correspondence $u_i \leftrightarrow v_i$, $1 \leq i \leq m$, induces a mapping of the set of messages onto the set of codes, under which each message $u_{i_1} \dots u_{i_k}$ is mapped into the code $v_{i_1} \dots v_{i_k}$. Since in a number of cases messages or codes may be decomposed into elementary components in more than one way, the indicated mapping, generally speaking, is not single-valued in either direction. Everywhere in what follows it is assumed that every message is decomposed into elementary messages in a unique way, and therefore the mapping of the set of messages onto the set of codes (encoding) is single-valued. A code system is said to have the property of **unique decipherability** if different messages are mapped into different codes.

We shall say that a code system has the property of **bounded delay** (cf. (4)) if it has the property of unique decipherability and there exists a number T such that, from the first T letters of any (not completely known to us) code, one can uniquely determine its first elementary code. If a code system has the property of bounded delay, then the least of the numbers T ($T \geq \max \lambda(v_i)$) satisfying the condition indicated above will be denoted by T_d and called the **deciphering delay**. The property of bounded delay makes it possible to carry out the deciphering of messages in a device with finite memory. From the first

Fig. 1

Figure 1: Fig. 1

T_d letters of any code $v_{i_1} v_{i_2} \dots v_{i_k}$ that have entered its memory, this device determines the first elementary code v_{i_1} and remembers the place from which the deciphering of the next code should begin; then, from the first T_d letters of the code $v_{i_2} \dots v_{i_k}$, it determines the elementary code v_{i_2} and remembers the place from which the deciphering of the next code should begin, and so on.

We shall say that a code system has the **synchronization property** if it has the property of unique ...

decoding, and there exists a number T such that every word of length T or more which is the beginning of some code and the end of some code forms a code. If a code system has the synchronization property, then the smallest of the numbers T ($T \geq 0$) satisfying the condition stated above will be denoted by T_s and will be called the **synchronization delay**. The meaning of these definitions is as follows. If the decoding device* at some moment in time begins to proceed from incorrect information about where the next code begins (a synchronization failure), **then, at most after $T_s + \max \lambda(v_i) - 1$ units of time, either it will start operating exactly as if this failure had not occurred (see Fig. 1*), or a word will enter its memory that is not the beginning of any code, which will signal the failure that has occurred.**

Fig. 1

2. Consider an arbitrary dictionary $V = \{v_1, \dots, v_m\}$ over some alphabet $B = \{b_1, \dots, b_r\}$. The dictionary $V^* = \{v_1^*, \dots, v_m^*\}$, where v_i^* , $1 \leq i \leq m$, denotes the reversal of the word v_i , will be called the **reversal of the dictionary V** , and the code system $\{A, U; B, V^*\}$ the **reversal of the code system $\{A, U; B, V\}$** . We shall call the dictionary V : 1) **free**; 2) **strongly free**; 3) **completely free**, if in the alphabet B there are no equalities of the form: 1) $v_{i_1} \dots v_{i_k} = v_{j_1} \dots v_{j_l}$, $i_1 \neq j_1$; 2) $v_{i_1} v_{i_2} \dots = v_{j_1} v_{j_2} \dots$, $i_1 \neq j_1$; 3) $\beta v_{i_1} \dots v_{i_k} = v_{j_1} \dots v_{j_l} \beta$, $0 < \lambda(\beta) < \lambda(v_{j_1})$, respectively. Denote by $K(V)$ (respectively $H(V)$) the set of all nonempty ends (beginnings) of words of the dictionary V which do not belong to V . It is easy to see that $H(V) = K^*(V^*)$ and $K(V) = H^*(V^*)$. Denote the number of elements of the set $K(V)$ by $N(V)$. Obviously, the number of elements of the set $H(V)$ is equal to $N(V^*)$.

Lemma 1. Let $v_{p_1}, \dots, v_{p_{m_0}}$ be all those words of the dictionary $V = \{v_1, \dots, v_m\}$ over $B = \{b_1, \dots, b_r\}$, each of which is not the end of any other word from V . Then

$$N(V) \leq \sum_{n=1}^{m_0} \lambda(v_{p_n}) - m_0 [\log_r m_0] - m + r \frac{r^{\lceil \log_r m_0 \rceil} - 1}{r - 1}, \quad (1)$$

where the estimate (1) is attained for any $m \geq m_0 \geq 1$, $r \geq 2$.

For the dictionary $V = \{v_1, \dots, v_m\}$ over B and an arbitrary set V_0 consisting of nonempty words in the alphabet B , construct the **right clas-**

* It will be shown below that a code system possessing the synchronization property has the property of bounded delay.

** This failure may occur, for example, as a result of a malfunction in the operation of the decoding device or an error in the transmission of the code.

*** In Fig. 1 it is shown that if a segment β of the incoming code $v_{i_1} v_{i_2} \dots v_{i_k}$ is the beginning of a code, the end of a code, and has length at least T_s , then, by the synchronization property, it forms some code $v_{j_1} v_{j_2} \dots v_{j_l}$. In this case the decoding device will operate exactly as if the code $v_{j_1} \dots v_{j_l} v'_{i_{k+1}} \dots v_{i_k}$ had entered its memory.

**** The reversal of the word $\beta = b_{i_1} b_{i_2} \dots b_{i_t}$ is the word $\beta^* = b_{i_t} b_{i_{t-1}} \dots b_{i_1}$.

***** Here $K^*(V^*) = (K(V^*))^*$, $H^*(V^*) = (H(V^*))^*$, etc.

with $R_n(V, V_0)$. As the initial class $R_0(V, V_0)$ take the set V_0 . The classes $R_n(V, V_0)$, $n = 1, 2, \dots$, will be defined successively as the sets of all nonempty words β in the alphabet B for which the equality $\beta_{n-1}\beta = v_i$ holds, or the equality $v_i\beta = \beta_{n-1}$, where $v_i \in V$, $\beta_{n-1} \in R_{n-1}(V, V_0)$. In an analogous way the left classes $L_n(V, V_0)$ are defined, and it is easy to show that $L_n(V, V_0) = R_n^*(V^*, V_0^*)$. We introduce the following notation:

$$R_n(V) = R_n(V, V), \quad L_n(V) = L_n(V, V), \quad \tilde{R}_n(V) = R_n(V, K(V)),$$

$$\tilde{L}_n(V) = L_n(V, H(V)).$$

Example 1. $A = U = \{a_1, a_2, a_3, a_4\}$, $B = \{0, 1\}$, $V = \{10, 101, 001, 0111\}$.

$R_0(V) =$ V	$R_1(V)$	$R_2(V)$	$R_3(V)$	$R_4(V)$	$R_5(V)$
10	1	0	01	11	
101		01	111		
001			11		
0111					

$\tilde{R}_0(V) =$ $K(V)$	$\tilde{R}_1(V)$	$\tilde{R}_2(V)$	$\tilde{R}_3(V)$	$\tilde{R}_4(V)$
0	01	11	11	
1	111	01		
01	0	111		

Fig. 2. Graph $G_R(V)$ for the code system of Example 1

Figure 2: Fig. 2. Graph $G_R(V)$ for the code system of Example 1

$\widetilde{R}_0(V) =$ $K(V)$	$\widetilde{R}_1(V)$	$\widetilde{R}_2(V)$	$\widetilde{R}_3(V)$	$\widetilde{R}_4(V)$
11	11			
111				

The principal auxiliary assertion is

Lemma 2. A word β belongs to $R_n(V, V_0)$, $n = 0, 1, \dots$, if and only if there exist words $\beta_0 \in V_0$, $v_{i_1}, \dots, v_{i_k}, v_{j_1}, \dots, v_{j_l} \in V$, $k + l = n$, such that the equality*

$$\beta_0 v_{i_1} \dots v_{i_k} \beta = v_{j_1} \dots v_{j_l}$$

holds, where $\lambda(\beta) < \lambda(v_{j_l})$, or the equality

$$\beta_0 v_{i_1} \dots v_{i_k} = v_{j_1} \dots v_{j_l} \beta,$$

where $\lambda(\beta) < \lambda(v_{i_k})$ (if $k \geq 1$); moreover, for no k' and l' , $0 \leq k' \leq k$, $0 \leq l' \leq l$, does the equality

$$\beta_0 v_{i_1} \dots v_{i_{k'}} = v_{j_1} \dots v_{j_{l'}}$$

hold.

- Let a code system $\{A, U; B, V\}$ be given (as was already said earlier, the dictionary U is assumed to be free). It is obvious that the code system has the property of unique decoding if and only if the dictionary V is free. On the other hand, it can be shown that the code system has the property of bounded delay (the synchronization property) if and only if the dictionary V is strongly free (respectively, completely free).

Fig. 2. Graph $G_R(V)$ for the code system of Example 1

Theorem 1.** In order that a code system have the property of unique decoding, it is necessary and sufficient that no class $R_n(V)$, for

$$1 \leq n \leq \min(N(V), N(V^*)) + 1$$

contain elementary codes.

Theorem 2. In order that a code system possessing the property of unique decoding have the property of bounded delay, it is necessary and sufficient that some class $R_n(V)$ with $n \leq N(V) + 1$ be empty; moreover, if it has the property of bounded delay, then

$$\left\lfloor \frac{n_1}{2} \right\rfloor \lambda_{\min} < T_d \leq \left\lceil \frac{n_1 + 1}{2} \right\rceil \lambda_{\max}, \quad (2)$$

where n_1 is the number of the first empty class $R_n(V)$,

$$\lambda_{\min} = \min_i \lambda(v_i), \quad \lambda_{\max} = \max_i \lambda(v_i).$$

Theorem 3. In order that a code system possessing the property of unique decoding have the synchronization property, it is necessary and sufficient that some class $\tilde{R}_n(V)$ with

* Here, as usual, the word $v_{i_1} \dots v_{i_k}$ is equal to the empty word if $k = 0$.

** The conditions of Theorem 1 were known earlier ⁽²⁾, but the question of the fact that, in order to verify these conditions, it is sufficient to examine a bounded number of classes $R_n(V)$, remained open.

$n \leq \min(N(V), N(V^*)) + 1$ was empty; moreover, if the code system has the synchronization property, then it and its inverse have the property of bounded delay, and for $\lambda_{\max} > 1$

$$\max(T_d, T_d^*) - 2(\lambda_{\max} - 1) \leq T_s \leq \left\lceil \frac{n_2}{2} \right\rceil \lambda_{\max}, \quad (3)$$

where n_2 is the number of the first empty class $\tilde{R}_n(V)$, and T_d^* is the decoding delay of the inverse code system.

Remark 1. Theorems 1 and 3 will remain valid if, in their formulations, the classes $R_n(V)$ and $\tilde{R}_n(V)$ are replaced by the classes $L_n(V)$ and $\tilde{L}_n(V)$, respectively*.

From Theorems 1 and 3, Remark 1, and the obvious equalities $R_n(V^*) = L_n^*(V)$, $\tilde{R}_n(V^*) = \tilde{L}_n^*(V)$, it follows that: 1) the inverse of a code system possessing the property of unique decodability possesses the property of unique decodability; 2) the inverse of a code system possessing the synchronization property possesses the synchronization property**.

4. There exists another, graphical method for recognizing the indicated properties of code systems, close to the method of Al. A. Markov ^(5,6), used by him for recognizing the first two of these properties. Construct, for the dictionary V , a directed graph $G_R(V)$. To do this, first take as the vertices of the graph the elements of the set $K(V) \cup V$, and connect the vertex β_i with the vertex β_j by an edge directed from β_i to β_j , if the equality $\beta_i \beta_j = v$ or the equality $\beta_i = v \beta_j$ holds, where $v \in V$, and then merge all vertices $\beta_i \in V$ into one common vertex and denote it by \bigwedge . In an analogous way (replacing the set $K(V)$ by $H(V)$ and changing the direction of the edges) the graph $G_L(V)$ is constructed.

Theorem 4. *In order that a code system possess any one of the following properties: 1) the property of unique decodability; 2) the property of bounded delay; 3) the synchronization property, it is necessary and sufficient that the graph $G_R(V)$ contain, respectively: 1) no directed cycles passing through the*

vertex \wedge ; 2) no directed cycles that can be reached from the vertex \wedge ; 3) no directed cycles at all.

With the aid of the graph $G_L(V)$ one can recognize the corresponding properties of the inverse code system, since the graphs $G_R(V^*)$ and $G_L(V)$ are isomorphic under the correspondence $\beta \leftrightarrow \beta^*$ between their vertices.

Received
22 V 1961

CITED LITERATURE

- ¹ D. A. Huffman, Proc. IRE, **40**, No. 9, 1098 (1952).
- ² A. A. Sardinas, G. W. Patterson, Conv. Rec. IRE, IT, part 8, 104 (1953).
- ³ M. P. Schützenberger, Trans. IRE, IT-2, No. 3, 47 (1956).
- ⁴ E. N. Gilbert, E. F. Moore, Bell Syst. Techn. J., **38**, No. 4, 933 (1959).
- ⁵ Al. A. Markov, DAN, **132**, No. 3, 521 (1960).
- ⁶ Al. A. Markov, DAN, **139**, No. 3 (1961).
- ⁷ S. W. Golomb, B. Gordon, L. R. Welch, Canad. J. Math., **10**, No. 2, 202 (1958).
- ⁸ E. N. Gilbert, Trans. IRE, IT-6, No. 4, 470 (1960).

* In terms of the classes $L_n(V)$ one can also answer a number of questions pertaining to other properties of code systems. In particular, it can be shown that (in the terminology of article (⁴)) an exhaustive code system is (for $\lambda_{\max} > 1$) completely self-synchronizing if and only if

$$H(V) = \bigcup_{n=1}^{n_3} L_n(V)$$

for some $n_3 \leq \frac{m-r}{r-1}$.

** The inverse of a code system possessing the property of bounded delay, generally speaking, does not have this property. In particular, the inverse of a completely self-synchronizing code system mentioned in the preceding footnote never has the property of bounded delay and, consequently, the latter does not have the synchronization property. Moreover, it can be shown that all other optimal Huffman code systems (¹) likewise do not have the synchronization property (for $m \geq 2r - 1$). Examples of code systems possessing the synchronization property are systems whose dictionary V contains a word without a comma (⁷) (in particular, a dictionary with a synchronizing prefix (⁸)).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.