



---

Soviet-era science, translated into English

# MATHEMATICS

L. N. KOROLEV

1957

SovietRxiv

---

View the original and related papers at <https://sovietrxiv.org/items/ru-195701.90255>

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

## Abstract

## Full Text

MATHEMATICS

L. N. KOROLEV

# CODING AND FOLDING OF CODES

(Presented by Academician M. A. Lavrent'ev, 25 X 1956)

§ 1. **General provisions.** By an **alphabet** we shall mean a finite set of elements. The elements of this set we shall call **letters**. The number of letters of an alphabet we shall call its **base**. We define a **word** of a given alphabet as an ordered finite set, to each element of which there is assigned one and only one letter of the alphabet; A. A. Markov uses a somewhat different definition of a word <sup>(1)</sup>. Words may contain identical letters and may consist of a single letter. Orderedness is the main point in the definition of a word. By the **length** of a word one should understand the number of all letters in the word, counting identical ones. The words of a given alphabet form the set of all words  $S$ .

If the alphabet is ordered in some way, i.e. if it is established which letter is to be considered the first letter of the alphabet, which the second, etc., then every word may be regarded as a positional representation of a certain positive number in a numeral system with base equal to the base of the alphabet.

Let the alphabet be ordered; this means that a one-to-one correspondence has been established between the letters of the alphabet and the numbers of the natural sequence from 0 to  $r - 1$ , where  $r$  is the base of the alphabet. Then to every word

$$a_{k_1} a_{k_2} \dots a_{k_n}$$

there is assigned, in a unique way, the number

$$k_1 r^{n-1} + k_2 r^{n-2} + \dots + k_n r^0 = \sum_{i=1}^n k_i r^{n-i}, \quad (1)$$

By virtue of the uniqueness of the representation of a number in the form (1), to each number there can be assigned one and only one word  $a_{k_1} a_{k_2} \dots a_{k_n}$ .

Alphabets with the same base we shall not distinguish and shall regard as equivalent.

Any mapping of the set of words of a given alphabet onto the set of words of the same or of another alphabet is called **coding**, and the images are called the **codes** of the corresponding words.

In many practical problems one considers not the whole set of words, but some part of this set, chosen according to definite laws. Therefore, in the subsequent discussion we shall consider a certain finite subset of the set of all words of a given alphabet, which we shall call a **dictionary**. By the **volume** of a dictionary we shall mean the number of words in the dictionary, and by the **length** of a dictionary—the sum of the lengths of all words composing it. Under coding of a dictionary various transformations of the lengths of individual words may occur, as well as of the length of the entire dictionary.

**Theorem 1.** *There exists a number  $d(r, N)$ , depending only on the base  $r$  of the coding alphabet and on the volume  $N$  of the dictionary, such that for any*

with one-to-one encoding, the length of the obtained dictionary  $l$  will be greater than or equal to  $d$ .

Thus, if under some encoding the length of the obtained dictionary  $l$  turns out to be smaller, then the encoding is ambiguous. It follows from Theorem 1 that under any method of one-to-one encoding the maximum length of a code in the dictionary is greater than or equal to  $E(\log_r N) + 1$ .

§ 2. **Contraction of codes.** Encoding algorithms may be divided into two types: alphabetic encoding and dictionary encoding.

**Alphabetic encoding** includes those algorithms for whose construction knowledge of the composition of the entire dictionary is not required. Alphabetic encoding may be applied to any word belonging to the set of all words, regardless of whether it belongs to the dictionary or not. Alphabetic encoding may be carried out, for example, by simply replacing each letter of the original word by the corresponding combination of letters of the encoding alphabet.

Since the dictionary is finite, another method of encoding is possible, which we shall call dictionary encoding. The dictionary is ordered, the ordinal numbers of its words are determined, and they are written in the symbols of the new alphabet. To construct the code of a word in this case it is necessary to know the entire dictionary, in order to be able to determine the ordinal number of this word in the dictionary. Dictionary encoding may be applied only to words belonging to the given dictionary.

Simple replacement is, generally speaking, irrational in terms of the lengths of the resulting codes.

The problem arises of reducing the length of codes, for example those obtained by simple replacement, or the problem of finding such an algorithm of alphabetic encoding which, for the words of a given dictionary, would yield codes of smaller length than simple replacement.

Let us define the operation  $\nabla_n$  of **contracting codes to length  $n$**  as an operation which assigns to every code of any length a code of length  $n$  over the same alphabet.

Examples of contraction operations may be the following: 1) contraction by discarding the letters of a word whose ordinal numbers are greater than  $n$ ; 2) contraction by deleting letters according to some criterion, for example the even ones; 3) division of the code of a word into parts, which are then added letterwise modulo  $r$ .

Contraction of codes may destroy the uniqueness of encoding. The application of a contraction operation to two different codes may give the same result. However, the following theorem is true:

**Theorem 2.** *There exist such classes of contraction operations  $\nabla_n$  for which the probability that the application of the operation  $\nabla_n$  to all codes of the dictionary will destroy the uniqueness of the encoding tends to zero as  $n$  increases.*

For the case of a dictionary that may be regarded as a random sample from the set of all words, an example of a contraction operation of such a class is an operation  $\nabla_n$  having the property that, after its application to a randomly chosen code, one should expect, with probability  $p = \frac{1}{r^n}$ , the appearance as the compressed code of any of the numbers from 0 to  $r^n - 1$ , where  $r$  is the base of the alphabet.

Operations of code contraction may be carried out on a computing machine. Code contraction makes it possible substantially to reduce the size of the storage device in computing machines, or to increase the amount of information transmitted over various communication channels.

Institute of Precision Mechanics and Computer Engineering  
Academy of Sciences of the USSR

Received  
10 X 1956

## CITED LITERATURE

1. A. A. Markov, *Tr. Matem. inst. im. V. A. Steklova AN SSSR*, **42** (1954).

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

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