Problems of Information Transmission, Vol. 36, No. 1, 2000, pp. 1–18. Translated from Problemy Peredachi Informatsii, No. 1, 2000, pp. 3–20.

INFORMATION THEORY

Pairs of Words with Nonmaterializable Mutual Information

A. E. Romashchenko

Received March 29, 1999; in final form, November 30, 1999

Abstract—Let there be a pair of words $\langle a, b\rangle$ with sufficiently large mutual information. Can we always "materialize" this information, i.e., point out a word c that can be computed from a and b simply and whose Kolmogorov complexity equals the mutual information between a and b? In this paper, we propose a better estimate for the amount of mutual information which may be materialized for words from the construction of Gács and Körner, and also give a new method for constructing pairs of words with nonmaterializable mutual information.

1. INTRODUCTION

Let there be two files x and y. Assume that some correlation exists between the data written in them. If we need to keep (or transmit) information about x and y in the most economical way, it is reasonable to demand that the "common" information of x and y should be kept in the form of a separate collection of data. The question arises whether it is possible to encode x by a pair of files $\langle u, v\rangle$ and encode y by a pair $\langle u, w\rangle$ in such a way that the length of u correspond to the amount of mutual information between the two source files. Then, instead of the pair $\langle x, y\rangle$, we could keep the triple $\langle u, v, w\rangle$. In doing so, we should be able to calculate u easily from x as well as from y.

Let us put this question in a more formal way. If we are given the words x and y, we can consider their Kolmogorov complexities and their mutual information. The mutual information $I(x : y)$ indicates how much the knowledge of one of these words simplifies the problem of generating the other:

$$I(x : y) := K(y) - K(y \,|\, x).$$

Sometimes, this quantity can be given a visual interpretation. Consider the simplest example. Let the word x be a concatenation of the words u and v and let the word y be a concatenation of the words u and w, namely,

$$x = uv, \qquad y = uw. \tag{1}$$

Assume that all three words u, v, and w are chosen as random and independent and their lengths are equal to n. Then the complexities of the words x and y are equal to $2n$, and their mutual information equals n. This fact is consistent with intuition: the words x and y have a common part u, and the mutual information is the complexity of this common part. The word u is a "materialization" of the mutual information between x and y.

In [1], the question was posed whether for any two words we can always find a third one, which materializes their mutual information. The word z materializing the mutual information between x and y should be easily calculable from each of them. Thus, we are interested in the question of if there always exists a word with a low complexity conditioned on each of the two given words, while its own complexity is equal to the mutual information between these words.

As is shown by Gács and Körner [1], this question is answered negatively. However, to formulate the exact statement, we need to explain what we mean by saying that some word z can easily be obtained from the word x and the word y (i.e., the Kolmogorov complexities $K(z \,|\, x)$ and $K(z \,|\, y)$

0032-9460/00/3601-0001 $25.00

are small). It seems to be impossible to define the relation of "conditional simplicity" for specific words (we cannot draw a boundary between "simple" and "complex" words). Therefore, let us turn from individual words to infinite word sequences and consider asymptotic properties of their complexities. In such terms, we can state the result of Gács and Körner.

Theorem 1. *There exist word sequences x_n and y_n such that*

$$K(x_n) = n + o(n), \qquad K(y_n) = n + o(n), \qquad I(x_n : y_n) = an + o(n)$$

(a is a positive constant), and for any word sequence z_n satisfying the condition

$$K(z_n \,|\, x_n) = o(n), \qquad K(z_n \,|\, y_n) = o(n),$$

it follows that $K(z_n) = o(n)$.

Speaking informally, this theorem asserts that there exist x_n and y_n such that if the complexity of z_n is small conditioned on x_n and on y_n, then the complexity of z_n on its own is also small. In addition, the mutual information of the words x_n and y_n grows linearly with n. Thus, there exist word sequences for which their mutual information cannot be materialized. Moreover, Theorem 1 claims that one cannot materialize even a part of the mutual information of the words x_n and y_n. More exactly, the amount of mutual information that can be materialized is infinitesimal with respect to n.

In [1], a certain class of examples of pairs $\langle x_n, y_n \rangle$ possessing the aforementioned property is described. In doing so, this construction permits us to generate such sequences x_n and y_n for any values of the parameter a, $0 < a < 1$; i.e., we can determine x_n and y_n, whose mutual information is very large (a is close to 1) but even a small part of it cannot be materialized.

No exact evaluation of the remainder terms was performed in [1]. But by analyzing the proof we can verify that Theorem 1 will still be true if we replace the terms $o(n)$ in its condition by $O(\sqrt{n})$ (or by $O(f(n))$, where $f(n)$ is any function growing faster than $\sqrt{n}$, but slower than n, i.e., $f(n) = o(n)$, $f(n) \geq \sqrt{n}$). However, in statements concerning the Kolmogorov complexity, it is natural to formulate equalities up to a logarithmic term. Indeed, such properties as the symmetry of mutual information

$$I(x : y) = I(y : x) + O(\log(|x| + |y|))$$

or the relation between the conditional complexity and the complexity of a pair of words

$$K(\langle x, y \rangle) = K(x) + K(y \,|\, x) + O(\log(|x| + |y|))$$

are valid up to the logarithm of the word length (see [2,3]). It therefore seems to be interesting to consider a strengthening of Theorem 1, namely, to prove it when $o(n)$ in its conditions is replaced by $O(\log n)$. More formally, a natural strengthening of Gács and Körner's theorem follows.

Theorem 2. *For any function $f(n)$ such that $f(n) = o(n)$ and $f(n) \geq \log n$, there exist word sequences x_n and y_n such that*

$$K(x_n) = n + O(f(n)), \qquad K(y_n) = n + O(f(n)), \qquad I(x_n : y_n) = an + O(f(n))$$

(a is some positive constant), and for any word sequence z_n satisfying the condition

$$K(z_n \,|\, x_n) = O(f(n)), \qquad K(z_n \,|\, y_n) = O(f(n)),$$

it follows that $K(z_n) = O(f(n))$.

In [4, 5], Theorem 2 was proved for an arbitrary value of the parameter a, $0 < a < 1$. Other examples of word sequence pairs with nonmaterializable mutual information (and, hence, other proofs of Theorem 2 for some special values of a) were also given in [6, 7].

Thus, for any $a < 1$, we can find words x_n and y_n such that their complexities are approximately equal to n, their mutual information is approximately equal to an, and their mutual information cannot be materialized. An. A. Muchnik has raised the question: What are the values of the parameter a such that for any x_n of complexity n it is possible to find y_n of complexity n with mutual information $I(x_n : y_n)$ which is approximately equal to an but cannot be materialized? More exactly, for what values of the parameter a is the following strengthening of Theorem 2 valid?

Theorem 3. *Let $f(n)$ be a function such that $f(n) = o(n)$ and $f(n) \geq \log n$, and let x_n be a sequence such that $K(x_n) = n + O(f(n))$. Then there exists a sequence y_n such that*

$$K(y_n) = n + O(f(n)), \qquad I(x_n : y_n) = an + O(f(n)),$$

and for any word sequence z_n satisfying the condition

$$K(z_n \mid x_n) = O(f(n)), \qquad K(z_n \mid y_n) = O(f(n)),$$

it follows that $K(z_n) = O(f(n))$.

For $a = 1/2$, this theorem was proved in [5].

In the present paper, we consider two arguments which make it possible to prove Theorem 3 for any values of the parameter a, $0 < a < 1$. The first of them utilizes an example of pairs of words given in [1]; the new method of the proof makes it possible to enhance the estimation of the amount of mutual information that can be materialized (and, thereby, to prove Theorems 2 and 3). The second argument is based on a new algebraic construction which generalizes the method of [5].

Remark 1. For the sake of simplicity, we prove Theorem 3 for $f(n) = \log n$ only. For the case of an arbitrary $f(n)$, all the arguments are carried over almost literally.

Notation Used

x, y, z, ... are binary words (finite sequences of zeroes and ones); the length of a word x is denoted by $|x|$;

$\boldsymbol{x} = \{x_n\}$, $\boldsymbol{y} = \{y_n\}$, $\boldsymbol{z} = \{z_n\}$, ... are infinite sequences of words; we assume that the word length in all sequences to be considered grows not faster than linearly: $|x_n| = O(n)$, $|y_n| = O(n)$, $|z_n| = O(n)$, ...; we say that the sequence $\boldsymbol{x} = \{x_n\}$ is simple conditioned on the sequence $\boldsymbol{y} = \{y_n\}$ if $K(x_n \mid y_n) = O(\log n)$;

$\langle x_1, x_2, \ldots, x_n \rangle$ is a tuple of binary words; we fix some computable enumeration of all finite tuples of words;

α, β, γ, ... are discrete random variables;

$K(x)$ is the Kolmogorov complexity of a word x;

$K(x_1, x_2, \ldots, x_n)$ is the Kolmogorov complexity of the index of a word tuple $\langle x_1, x_2, \ldots, x_n \rangle$ in the chosen enumeration;

$K(x \mid y)$ is the Kolmogorov complexity of a word x conditioned on a word y;

$K(x_1, x_2, \ldots, x_n \mid y_1, y_2, \ldots, y_m)$ is the Kolmogorov complexity of the index of a tuple $\langle x_1, x_2, \ldots, x_n \rangle$ conditioned on the index of a tuple $\langle y_1, y_2, \ldots, y_m \rangle$;

$I(x : y) := K(y) - K(y \mid x)$ is the mutual information of words x and y;

$I(x : y \mid z) := K(y \mid z) - K(y \mid x, z)$ is the mutual information of words x and y conditioned on z; the mutual information of word tuples (conditioned on word tuples) is defined analogously;

$H(\alpha)$ is Shannon's entropy of a random variable α;

$\mathcal{I}(\alpha:\beta) := H(\alpha) + H(\beta) - H(\alpha,\beta)$ is the mutual information of random variables α and β;

let $\langle x_1, x_2, \ldots, x_k\rangle$ be a word tuple and let $V = \{i_1, i_2, \ldots, i_r\} \subseteq \{1,2,\ldots,k\}$ be a set of indices; then we denote by x^V the tuple $\langle x_{i_1}, x_{i_2}, \ldots, x_{i_r}\rangle$; a similar designation is used for tuples of random variables;

in the paper, all logarithms are taken to the base 2.

2. STOCHASTIC PAIRS

Let us consider a particular case from the family of Gács and Körner's examples, for which we give a new simpler proof of nonmaterializability of the mutual information. The new argument will allow us to improve the estimate for the amount of information separated and obtain proofs for Theorems 2 and 3.

Before we begin the proof, let us give several definitions.

Definition 1. Let random variables $\varphi^1, \varphi^2, \ldots, \varphi^k$ assume values in finite alphabets $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_k$ and possess the joint distribution

$$P(a^1, a^2, \ldots, a^k) = \mathrm{Prob}[\varphi^1 = a^1, \varphi^2 = a^2, \ldots, \varphi^k = a^k].$$

Then the tuple of infinite word sequences $\langle \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^k\rangle$ is called *P-typical* if for any set of values $\langle a^1, a^2, \ldots, a^k\rangle$, the number of positions i such that in every word x_n^j the ith place is occupied by the letter a^j is equal to $nP(a^1, a^2, \ldots, a^k) + O(1)$.

Definition 2. We call a P-typical tuple of word sequences $\langle \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^k\rangle$ *P-random* if for any nonempty set of indices $V \subseteq \{1,2,\ldots,k\}$ and any (maybe, empty) set of indices $W \subseteq \{1,2,\ldots,k\}$, the equality

$$K(x_n^V \mid x_n^W) = nH(\varphi^V \mid \varphi^W) + O(\log n) \tag{2}$$

holds. (In the case where the set of indices W is empty, the conditional Kolmogorov complexity and conditional Shannon entropy in (2) become unconditional.)

Remark 2. Let a pair of sequences $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ be random with respect to the distribution P of a pair of random variables $\langle \varphi, \psi\rangle$. Then the values of $K(x_n)$, $K(y_n)$, and $K(x_n, y_n)$ are equal (up to a logarithmic addend) to $nH(\varphi)$, $nH(\psi)$, and $nH(\varphi,\psi)$, respectively. Hence, the analogous equality is valid for the mutual information as well. Indeed,

$$\mathcal{I}(\varphi:\psi) = H(\varphi) + H(\psi) - H(\varphi,\psi)$$

and

$$I(x_n : y_n) = K(x_n) + K(y_n) - K(x_n, y_n) + O(\log n).$$

Therefore,

$$I(x_n : y_n) = n\mathcal{I}(\varphi:\psi) + O(\log n).$$

Proposition 1. (1) *If random variables $\varphi^1, \varphi^2, \ldots, \varphi^k$ possess a joint distribution P, then for any P-typical sequences $\boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^k$, any nonempty set of indices $V \subseteq \{1,2,\ldots,k\}$, and any (maybe, empty) set of indices $W \subseteq \{1,2,\ldots,k\}$, we have the inequality*

$$K(x_n^V \mid x_n^W) \le nH(\varphi^V \mid \varphi^W) + O(\log n).$$

(2) *For any distribution P, there exist P-random tuples.*

Proof. For $k = 1$, the proof is given in [2], as well as in [3]. In the general case, the proof is similar. $\triangle$

Remark 3. Let a pair of sequences $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ be typical with respect to the distribution P of a pair of random variables $\langle \varphi, \psi \rangle$, and let $K(x_n, y_n) = nH(\varphi, \psi) + O(\log n)$. Then the given pair is P-random. Let us prove, for example, that $K(x_n) = nH(\varphi) + O(\log n)$. Indeed, due to Proposition 1, we have the inequality $K(x_n) \le nH(\varphi) + O(\log n)$. To prove the converse, it suffices to add together the equalities

$$K(x_n, y_n) = K(x_n) + K(y_n \,|\, x_n) + O(\log n),$$
$$nH(\varphi, \psi) = K(x_n, y_n) + O(\log n)$$

and the inequality

$$K(y_n \,|\, x_n) \le nH(\psi \,|\, \varphi) + O(\log n).$$

In a similar way, we can prove the following equalities:

$$K(y_n) = nH(\psi) + O(\log n),$$
$$K(x_n \,|\, y_n) = nH(\varphi \,|\, \psi) + O(\log n),$$
$$K(y_n \,|\, x_n) = nH(\psi \,|\, \varphi) + O(\log n).$$

In what follows, we need a construction, which will enable us to extend P-random tuples. Specifically, let k and l be positive integers with $l < k$. We assume that random variables $\varphi^1, \varphi^2, \ldots, \varphi^k$ have a joint distribution P. We denote by P' the projection of P on the first l coordinates, i.e., the joint distribution of $\varphi^1, \varphi^2, \ldots, \varphi^l$. Then any P'-random tuple can be complemented to a P-random one. More exactly, we have the following lemma.

Lemma 1. *Assume that P is the joint distribution of k random variables, each of which takes two values 0 or 1, P' is the projection of the distribution P on the first l coordinates, and a tuple $\langle \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^l \rangle$ is P'-random. Then there exist sequences $\boldsymbol{x}^{l+1}, \boldsymbol{x}^{l+2}, \ldots, \boldsymbol{x}^k$ such that the tuple $\langle \boldsymbol{x}^1, \boldsymbol{x}^2, \ldots, \boldsymbol{x}^k \rangle$ is P-random.*

Proof. Let us give the proof for the case of $k = 2$, $l = 1$ (the proof in the general case is analogous).

Let binary random variables φ, ψ have a joint distribution P:

$$\mathrm{Prob}[\varphi = i, \psi = j] = p_{ij},$$

where $i, j = 0, 1$. In the case considered, P' is the distribution of φ:

$$\mathrm{Prob}[\varphi = i] = p_{i0} + p_{i1}.$$

Assume that we are given a P'-random word sequence $\boldsymbol{x}$. We need to find a sequence $\boldsymbol{y}$ such that $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ is P-random.

Let the characters 0 and 1 occur s_0 and s_1 times, respectively, in the word x_n (the values of s_0 and s_1 depend on n). Since the sequence $\boldsymbol{x}$ is P'-random, we have

$$s_i = n(p_{i0} + p_{i1}) + O(1)$$

for $i = 0, 1$.

Consequently, we can represent the numbers s_0 and s_1 as sums $s_i = s_{i0} + s_{i1}$ so that

$$s_{ij} = np_{ij} + O(1)$$

(the numbers s_{ij} will also depend on n).

We call a binary word $\widehat{y}$ of a length n admissible if it can be obtained from the word x_n by the following transformation: in the word x_n, we must replace $(s_{01} + s_{10})$ bits by their complements, namely, replace s_{01} zeroes by ones, and s_{10} ones by zeroes.

Clearly, if $\widehat{y}$ is admissible, then the frequencies of ones and zeroes in this word correspond to the probability distribution ψ, and the pair $\langle x_n, \widehat{y}\rangle$ is P-typical. It remains to choose an admissible $\widehat{y}$ such that the resulting pair is P-random. But for this purpose, it suffices to take an admissible word having the largest complexity conditioned on x_n.

For a fixed x_n, the number of all admissible words $\widehat{y}$ is equal to $(C_{s_0}^{s_{01}} \cdot C_{s_1}^{s_{10}})$. If, as y_n, we take an admissible word of the maximum complexity conditioned on x_n, then

$$K(\widehat{y}_n \,|\, x_n) = \log(C_{s_0}^{s_{01}} \cdot C_{s_1}^{s_{10}}) + O(\log n) = nH(\psi \,|\, \varphi) + O(\log n).$$

(The latter equality can easily be proved by estimating the values of the binomial coefficients with Stirling's formula.) Furthermore, since the sequence $\boldsymbol{x}$ is P-random, $K(x_n) = nH(\varphi) + O(\log n)$. By the equality

$$K(x_n, y_n) = K(x_n) + K(y_n \,|\, x_n) + O(\log n),$$

we obtain

$$K(x_n, y_n) = nH(\varphi) + nH(\psi \,|\, \varphi) + O(\log n) = nH(\varphi, \psi) + O(\log n). \tag{3}$$

According to Remark 3, from (3) it follows that the pair constructed is P-random. $\triangle$

Lemma 1 is valid not only for binary but also for arbitrary jointly distributed random variables. However, we have proved (and will use) it only for joint distributions of binary random variables.

Let us introduce one more notion. Consider the joint distributions of a pair of random variables $\langle \varphi, \psi\rangle$ with the following properties: both variables φ, ψ take values 0 and 1 with probability 1/2 (i.e., they are binary uniformly distributed); in addition, φ and ψ take different values with probability α (and, accordingly, coincide with probability $(1 - \alpha)$). Thus,

$$\mathrm{Prob}[\varphi = 0, \psi = 0] = \mathrm{Prob}[\varphi = 1, \psi = 1] = \frac{1-\alpha}{2},$$
$$\mathrm{Prob}[\varphi = 1, \psi = 0] = \mathrm{Prob}[\varphi = 0, \psi = 1] = \frac{\alpha}{2}.$$

This distribution is given by Table 1.

Table 1. Distribution P of the pair of random variables φ, ψ

φ \ ψ	0	1
0	$\frac{1-\alpha}{2}$	$\frac{\alpha}{2}$
1	$\frac{\alpha}{2}$	$\frac{1-\alpha}{2}$

Definition 3. Let us call word sequences $\boldsymbol{x}$, $\boldsymbol{y}$ an α-pair if they are a P-random pair with respect to the distribution P given by Table 1.

Let random variables φ and ψ have a joint distribution P specified in Table 1, and let sequences $\boldsymbol{x}$ and $\boldsymbol{y}$ form an α-pair. Then x_n, y_n are random words of length n and differ from each other at $\alpha n + O(1)$ positions.

Furthermore, one can easily calculate Shannon's entropies of the random variables φ and ψ, as well as the entropy of the pair $\langle\varphi, \psi\rangle$,

$$H(\varphi) = H(\psi) = 1, \qquad H(\varphi, \psi) = 1 - (\alpha \log \alpha + (1-\alpha)\log(1-\alpha)).$$

We obtain the mutual information of the given random variables

$$\mathcal{I}(\varphi : \psi) = 1 + (\alpha \log \alpha + (1-\alpha)\log(1-\alpha)).$$

Denote the amount of mutual information

$$c(\alpha) = 1 + (\alpha \log \alpha + (1-\alpha)\log(1-\alpha)). \tag{4}$$

Obviously, $c(\alpha) > 0$ for $\alpha \neq 1/2$.

If $\boldsymbol{x}$ and $\boldsymbol{y}$ form an α-pair, then

$$\begin{aligned} K(x_n) &= n + O(\log n), \\ K(y_n) &= n + O(\log n), \\ I(x_n, y_n) &= c(\alpha)n + O(\log n). \end{aligned}$$

Thus, for $\alpha \neq 1/2$ the mutual information of x_n and y_n grows linearly in n.

The case $\alpha = 1/2$ should be considered separately. Since $c(1/2) = 0$, the mutual information between the words x_n and y_n is equal to $O(\log n)$. This agrees with intuition: if two words are chosen randomly and independently, they differ approximately in a half of bits. But according to the definition of a 1/2-pair, the words x_n and y_n must actually be a pair of random words which differ approximately in a half of bits.

Exactly α-pairs appear to be examples of words satisfying Theorem 2. We will prove that for any $\alpha \in (0; 1)$ the mutual information of random α-pairs cannot be materialized. To this end, we need the following technical lemmas.

Lemma 2. *Let a sequence $\boldsymbol{z}$ be simple conditioned on $\boldsymbol{x}$ and on $\boldsymbol{y}$, i.e., $K(z_n \,|\, x_n) = O(\log n)$ and $K(z_n \,|\, y_n) = O(\log n)$. Then $K(z_n) \leq I(x_n : y_n) + O(\log n)$.*

Proof. For any x_n, y_n, z_n we have the relations

$$\begin{aligned} K(x_n, z_n) &= K(x_n) + K(z_n \,|\, x_n) + O(\log n), \\ K(x_n) - K(x_n \,|\, y_n) &= I(x_n : y_n) + O(\log n), \\ K(x_n \,|\, y_n) &\leq K(x_n \,|\, z_n) + K(z_n \,|\, y_n) + O(\log n), \\ K(x_n \,|\, z_n) + K(z_n) &= K(x_n, z_n) + O(\log n). \end{aligned}$$

By adding them together, we obtain

$$K(z_n) \leq K(z_n \,|\, x_n) + K(z_n \,|\, y_n) + I(x_n : y_n) + O(\log n). \tag{5}$$

Taking into account that $K(z_n \,|\, x_n) = O(\log n)$ and $K(z_n \,|\, y_n) = O(\log n)$, we obtain $K(z_n) \leq I(x_n : y_n) + O(\log n)$. $\triangle$

Lemma 3. *Assume that we are given four sequences $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{x}'$, $\boldsymbol{y}'$ such that x_n and y_n are independent conditioned on x'_n and on y'_n:*

$$I(x_n : y_n \,|\, x'_n) = O(\log n), \qquad I(x_n : y_n \,|\, y'_n) = O(\log n).$$

Then every sequence $\boldsymbol{z}$, which is simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$ (so that $K(z_n \,|\, x_n) = O(\log n)$ and $K(z_n \,|\, y_n) = O(\log n)$), is also simple conditioned on $\boldsymbol{x}'$ and $\boldsymbol{y}'$ (so that $K(z_n \,|\, x'_n) = O(\log n)$ and $K(z_n \,|\, y'_n) = O(\log n)$).

Proof. Let $\boldsymbol{z}$ be simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$. Let us show that $\boldsymbol{z}$ is also simple conditioned on $\boldsymbol{x}'$ and $\boldsymbol{y}'$. Consider a relativized version of the inequality (5)

$$K(z_n \mid x'_n) \le K(z_n \mid x_n, x'_n) + K(z_n \mid y_n, x'_n) + I(x_n : y_n \mid x'_n) + O(\log n).$$

By weakening it, we obtain

$$K(z_n \mid x'_n) \le K(z_n \mid x_n) + K(z_n \mid y_n) + I(x_n : y_n \mid x'_n) + O(\log n).$$

Since $\boldsymbol{z}$ is simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$ and the sequences $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent conditioned on $\boldsymbol{x}'$, we obtain $K(z_n \mid x'_n) = O(\log n)$. Similarly, $K(z_n \mid y'_n) = O(\log n)$. $\triangle$

The following lemma will allow us to reduce the problem of nonmaterializability of the mutual information for an α-pair to that for some β-pair, where $\beta > \alpha$.

Lemma 4. *Let $\alpha < 1/2$ and let the sequences $\boldsymbol{x}$ and $\boldsymbol{y}$ be an α-pair. Then for any β such that*

$$\alpha < \beta \le \min\{1 - \sqrt{1-2\alpha}, 1/2\},$$

a β-pair $\boldsymbol{x}'$, $\boldsymbol{y}'$ exists such that any sequence $\boldsymbol{z}$, which is simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$, is also simple conditioned on $\boldsymbol{x}'$ and $\boldsymbol{y}'$.

Proof. Let us construct sequences $\boldsymbol{x}'$, $\boldsymbol{y}'$, which form a β-pair and are such that

$$I(x_n : y_n \mid x'_n) = O(\log n), \qquad I(x_n : y_n \mid y'_n) = O(\log n).$$

Then, according to Lemma 3, every sequence, which is simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$, is also simple conditioned on $\boldsymbol{x}'$ and $\boldsymbol{y}'$. For constructing $\boldsymbol{x}'$ and $\boldsymbol{y}'$, let us use the properties of quadruples of sequences of a special form.

Consider a quadruple of binary random variables φ_1, φ_2, φ_3, and φ_4, which have the following joint distribution P'. First of all, assume that the joint distributions of the pairs of these random variables $\langle\varphi_1, \varphi_2\rangle$ and $\langle\varphi_3, \varphi_4\rangle$ are as shown in Table 2. Secondly, the conditional probability

Table 2. Projections of distribution P'

Distribution of φ_1 and φ_2			Distribution of φ_3 and φ_4		
$\varphi_1 \backslash \varphi_2$	0	1	$\varphi_3 \backslash \varphi_4$	0	1
0	$\frac{1-\alpha}{2}$	$\frac{\alpha}{2}$	0	$\frac{1-\beta}{2}$	$\frac{\beta}{2}$
1	$\frac{\alpha}{2}$	$\frac{1-\alpha}{2}$	1	$\frac{\beta}{2}$	$\frac{1-\beta}{2}$

distributions of the pair $\langle\varphi_1, \varphi_2\rangle$ for known values of φ_3 and φ_4 must be such as in Table 3. In doing so, let us take $\dfrac{1-\beta-\sqrt{1-2\alpha}}{2}$ as the value of the parameter t. The expression for the quantity t makes sense (the radicand is nonnegative) since by definition $\alpha \le 1/2$.

The given definition is correct, i.e., Tables 2 and 3 actually specify a joint distribution of a quadruple of random variables if all numbers in the tables are nonnegative. This condition holds if $t \ge 0$ and $1 - \beta - t \ge 0$. It is easy to verify that both inequalities are satisfied for the chosen value of the parameter. Indeed, the first inequality follows from the restriction $\beta \le 1 - \sqrt{1-2\alpha}$ in the condition of the lemma and the second inequality is valid for any α and β from the interval $(0, 1/2)$.

Table 3. Distributions of φ_1, φ_2 for fixed values of φ_3, φ_4

$\varphi_3 = 0,\ \varphi_4 = 0$		
$\varphi_1 \backslash \varphi_2$	0	1
0	$\frac{1-\beta-t}{1-\beta}$	0
1	0	$\frac{t}{1-\beta}$

$\varphi_3 = 0,\ \varphi_4 = 1$		
$\varphi_1 \backslash \varphi_2$	0	1
0	$\frac{\beta-\alpha}{2\beta}$	$\frac{\alpha}{2\beta}$
1	$\frac{\alpha}{2\beta}$	$\frac{\beta-\alpha}{2\beta}$

$\varphi_3 = 1,\ \varphi_4 = 0$		
$\varphi_1 \backslash \varphi_2$	0	1
0	$\frac{\beta-\alpha}{2\beta}$	$\frac{\alpha}{2\beta}$
1	$\frac{\alpha}{2\beta}$	$\frac{\beta-\alpha}{2\beta}$

$\varphi_3 = 1,\ \varphi_4 = 1$		
$\varphi_1 \backslash \varphi_2$	0	1
0	$\frac{t}{1-\beta}$	0
1	0	$\frac{1-\beta-t}{1-\beta}$

Let us prove that for the chosen value of t, the random variables φ_1 and φ_2 are independent conditioned on φ_3 and φ_4, and hence

$$\mathcal{I}(\varphi_1 : \varphi_2 \,|\, \varphi_3) = 0, \qquad \mathcal{I}(\varphi_1 : \varphi_2 \,|\, \varphi_4) = 0.$$

The proofs of the independence of φ_1 and φ_2 conditioned on $\varphi_3 = 0$, $\varphi_4 = 0$, $\varphi_3 = 1$, and $\varphi_4 = 1$ are completely analogous, and we consider only the case of $\varphi_3 = 0$.

So let us prove that if we choose the parameter t as indicated, the random variables φ_1 and φ_2 are independent conditioned on $\varphi_3 = 0$. One can easily see that the joint distribution of φ_1 and φ_2 conditioned on $\varphi_3 = 0$ is such as shown in Table 4. Independence of a pair of binary random

Table 4. Distribution of φ_1 and φ_2 conditioned on $\varphi_3 = 0$

$\varphi_1 \backslash \varphi_2$	0	1
0	$1-\beta-t+\frac{\beta-\alpha}{2}$	$\frac{\alpha}{2}$
1	$\frac{\alpha}{2}$	$t+\frac{\beta-\alpha}{2}$

matrices means that the distribution-generating matrix of four numbers is of rank one. It remains to find a value t such that the determinant of the matrix given by Table 4 equals zero. We obtain the quadratic equation

$$\left(1-\beta-t+\frac{\beta-\alpha}{2}\right)\left(t+\frac{\beta-\alpha}{2}\right)-\frac{\alpha^2}{4}=0,$$

one of whose roots is exactly the number $\dfrac{1-\beta-\sqrt{1-2\alpha}}{2}$.

We have proved that the random variables φ_1 and φ_2 are independent conditioned on φ_3 and φ_4. Consequently, if the sequences $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{y}'$ form a P-random quadruple, then

$$I(x_n : y_n \,|\, x'_n) = O(\log n), \qquad I(x_n : y_n \,|\, y'_n) = O(\log n).$$

Moreover, $\boldsymbol{x}$ and $\boldsymbol{y}$ form an α-pair, and $\boldsymbol{x}'$ and $\boldsymbol{y}'$ form a β-pair.

To prove the lemma, it remains to note that if we are given an arbitrary α-pair $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$, then by Lemma 1 it can be completed to a P-random quadruple $\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{y}' \rangle$. $\triangle$

Corollary 1. *Let $\frac{3}{8} < \alpha < \frac{1}{2}$, and let $\boldsymbol{x}$, $\boldsymbol{y}$ be an α-pair. Then any sequence $\boldsymbol{z}$ which is simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$ has a logarithmic complexity, i.e., $K(z_n) = O(\log n)$.*

Proof. Let $\boldsymbol{z}$ be a sequence simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$. It can easily be verified that $\frac{1}{2} < 1 - \sqrt{1-2\alpha}$, i.e., $\alpha \in (3/8, 1/2)$ and $\beta = 1/2$ satisfy the condition of Lemma 4. Therefore, $\boldsymbol{z}$ is also simple conditioned on some $\boldsymbol{x}'$ and $\boldsymbol{y}'$, which form a 1/2-pair. But the mutual information between x'_n and y'_n does not exceed $O(\log n)$. By Lemma 2, we obtain $K(z_n) = O(\log n)$. $\triangle$

Now let α be an arbitrary number from the interval $(0, 1)$. In order to prove that for any α-pair the mutual information is nonmaterializable, it suffices to repeat the technique from the proof of Corollary 1 several times. Let us state this in a formal way.

Proposition 2. *Let $0 < \alpha < 1$, and let $\boldsymbol{x}$ and $\boldsymbol{y}$ be an α-pair. Then for every sequence $\boldsymbol{z}$ which is simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$,*

$$K(z_n) = O(\log n).$$

Proof. So, let $\boldsymbol{z}$ be simple conditioned on $\boldsymbol{x}$ and on $\boldsymbol{y}$. First of all, let us note that if $\boldsymbol{x}$ and $\boldsymbol{y}$ are an α-pair, then, by replacing all bits of words from the sequence $\boldsymbol{y}$ by their complements, we obtain a $(1/2 - \alpha)$-pair with the same properties of materializability of the mutual information. Therefore, it suffices to consider $\alpha \le 1/2$.

The case $\alpha = 1/2$ is trivial. The words of a random 1/2-pair are independent, i.e., $I(x_n : y_n) = O(\log n)$. By Lemma 2, the complexity of the word z_n is not greater than the mutual information of x_n and y_n. So it only remains to consider $\alpha < 1/2$.

Now let $0 < \alpha < 1/2$. Let us choose the parameter α^1 as

$$\alpha^1 = \min\{1 - \sqrt{1-2\alpha}, 1/2\}.$$

By Lemma 3, there exists an α^1-pair $\boldsymbol{x}^1$, $\boldsymbol{y}^1$ such that any sequence $\boldsymbol{z}$ simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$ is simple conditioned on $\boldsymbol{x}^1$, $\boldsymbol{y}^1$ too. If $\alpha_1 = 1/2$, the proof is complete. Otherwise, let us again apply Lemma 3, according to which there exists an α^2-pair $\boldsymbol{x}^2$, $\boldsymbol{y}^2$ with

$$\alpha^2 = \min\{1 - \sqrt{1-2\alpha^1}, 1/2\},$$

such that every $\boldsymbol{z}$ simple conditioned on $\boldsymbol{x}^1$ and $\boldsymbol{y}^1$ is also simple conditioned on $\boldsymbol{x}^2, \boldsymbol{y}^2$. By applying Lemma 3 repeatedly, we obtain

$$\langle \boldsymbol{x}^1, \boldsymbol{y}^1 \rangle, \langle \boldsymbol{x}^2, \boldsymbol{y}^2 \rangle, \ldots, \langle \boldsymbol{x}^n, \boldsymbol{y}^n \rangle, \ldots,$$

where for any n the sequences $\boldsymbol{x}^n$ and $\boldsymbol{y}^n$ form an α^n-pair,

$$\alpha^{n+1} = \min\{1 - \sqrt{1-2\alpha^n}, 1/2\}, \quad n = 1, 2, \ldots. \tag{6}$$

At the same time, every sequence $\boldsymbol{z}$ simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$ is also simple conditioned on each of the sequences $\boldsymbol{x}^n$ and $\boldsymbol{y}^n$. It remains to prove that at some step we will obtain a 1/2-pair, i.e.,

$$\exists N \quad \alpha^N = 1/2.$$

Assume the contrary. Then $\{\alpha^n\}$ is an infinite strictly increasing sequence, all members of which are less than 1/2. Therefore, the sequence converges to some limit α_∞. Substituting α_∞ into the recurrent relation (6), we obtain

$$\alpha_\infty = 1 - \sqrt{1-2\alpha_\infty},$$

whence $\alpha_\infty = 0$. But this contradicts the increasing of α_n. $\triangle$

Proposition 2 make it possible to prove Theorem 2 for any value of the parameter a. Indeed, according to (4), for any a from the interval $(0,1)$ there exists α such that $c(\alpha) = a$; then, for an α-pair $\boldsymbol{x}$, $\boldsymbol{y}$,

$$K(x_n) = n + O(\log n), \qquad K(y_n) = n + O(\log n), \qquad I(x_n : y_n) = an + O(\log n).$$

But by Proposition 2, for any α-pair the mutual information is nonmaterializable. To prove Theorem 3, it remains to note that if $\alpha \in (0,1)$ and $\boldsymbol{x}$ is a sequence such that $|x_n| = n$ and $K(x_n) = n + O(\log n)$, then by Lemma 1 it is possible to find a sequence $\boldsymbol{y}$ such that $\boldsymbol{x}$ and $\boldsymbol{y}$ form an α-pair.

3. ORTHOGONAL LINEAR SUBSPACES

In this section, we consider a second construction which allows us to obtain word sequences $\boldsymbol{x}$, $\boldsymbol{y}$ with nonmaterializable mutual information. Let us fix two parameters, namely, positive integers m and k such that $2k < m$. For any $n \in \mathbb{N}$, let us choose a finite field F_n (the field F_n consists of $2^{\Theta(n)}$ elements). Now we denote by V_n the m-dimensional linear space over F_n. We assume that some basis is fixed in V_n. We say that vectors v and w from V_n are *orthogonal* if in the basis fixed,

$$v = (v^1, v^2, \ldots, v^m), \qquad w = (w^1, w^2, \ldots, w^m),$$

and

$$v^1 w^1 + v^2 w^2 + \ldots + v^m w^m = 0.$$

Accordingly, we call linear subspaces $A, B \subseteq V_n$ *orthogonal* if any vector from A is orthogonal to every vector from B.

As x_n and y_n, we take pairs of orthogonal k-dimensional subspaces of V_n. If P_n is the number of all pairs of orthogonal k-dimensional subspaces in V_n, then, obviously, $K(x_n, y_n) \le \log P_n + O(\log n)$. We are interested in random pairs $\langle x_n, y_n \rangle$, i.e., pairs such that $K(x_n, y_n) = \log P_n + O(\log n)$. Note that for $k = 1$ and $m = 3$ we obtain the construction from [5].

The parameters of the given construction are the numbers m and k, as well as the sizes of the fields F_n. Below we will choose values of the parameters such that the complexities of x_n and y_n will be close to n, and the quantity $I(x_n : y_n)$ will depend on the ratio of k and m. Most interesting is the case where k is chosen close to $m/2$ since the mutual information between x_n and y_n appears then to be close to n.

For fixed values of the parameters, any word x_n of complexity n may be considered as a code in a random k-dimensional linear subspace of V_n. For any k-dimensional subspace x_n, we can find a k-dimensional subspace y_n orthogonal to it such that the conditional complexity $K(y_n \mid x_n)$ has the maximum possible magnitude (more exactly, $K(y_n \mid x_n)$ is the logarithm of the number of k-dimensional subspaces in V_n that are orthogonal to the subspace x_n). It is clear that the pair $\langle x_n, y_n \rangle$ will then be random, i.e., will have a complexity of $\log P_n + O(\log n)$. To obtain the proof of Theorem 3, it remains to verify that Theorem 2 holds for random pairs of orthogonal subspaces (the mutual information cannot be materialized).

The proof is based on the following property of orthogonal subspaces. Consider a graph G_n, whose vertices are all k-dimensional subspaces in V_n. Edges in this graph connect orthogonal subspaces. Fix some vertex v_0 of the graph. Consider a random walk along this graph starting in v_0. Let v_i be the graph vertex where the random walk appears after i steps. For any i, the random variable v_i is distributed on the vertex set of G_n (i.e., on the set of k-dimensional subspaces of V_n). Let us show that for some s the distribution of v_s is close to the uniform one. This s depends on m and k, but does not depend on n.

So, for any n, let the words x_n and y_n encode a random pair of orthogonal k-dimensional subspaces of V_n. The formal proof is as follows.

Proposition 3. *The Kolmogorov complexities and mutual information of* $\boldsymbol{x}$, $\boldsymbol{y}$ *grow linearly in* n, *namely,*

$$K(x_n) = (mk - k^2)|F_n| + O(\log n), \tag{7}$$

$$K(y_n) = (mk - k^2)|F_n| + O(\log n), \tag{8}$$

$$I(x_n : y_n) = k^2|F_n| + O(\log n). \tag{9}$$

Proof. Let W be a linear space over F_n. Let us find the number of sequences $e_1, e_2, \ldots, e_k$ consisting of k linearly independent vectors of the space W.

Denote $s = \dim(W)$ and $N = |F_n|$. As e_1, we can take any nonzero vector of W. Thus, to choose the first vector in a sequence, we have $N^s - 1$ possibilities. Let a vector e_1 be already chosen. Then, to choose the second vector in a sequence, we have $(N^s - N)$ possibilities since e_2 must be linearly independent of e_1. Next, if we have chosen the first i vectors of a sequence, then, as a vector e_{i+1}, we can take any vector, which does not belong to the linear span of $e_1, \ldots, e_i$; i.e., to choose e_{i+1}, we have $N^s - N^i$ possibilities. Consequently, in the space W we have

$$(N^s - 1)(N^s - N)\ldots(N^s - N^{k-1}) = N^{ks}(1 + O(1/N))$$

sequences of k linearly independent vectors. Replacing s with the number m, we find the number of sequences of k linearly independent vectors in the whole space V_n. Furthermore, substituting the number k instead of s, we find the number of sequences from k linearly independent vectors in every k-dimensional subspace of V_n. The ratio of these quantities,

$$Q_n = N^{mk-k^2}(1 + O(1/N)),$$

gives the number of k-dimensional subspaces of V_n. Since x_n and y_n are chosen at random,

$$K(x_n) = \log Q_n + O(\log n), \qquad K(y_n) = \log Q_n + O(\log n).$$

It remains to find the mutual information between x_n and y_n.

Let us calculate the magnitude of the conditional complexity $K(y_n \,|\, x_n)$. If the subspace x_n is already fixed, then y_n lies in the subspace of vectors of V_n orthogonal to x_n. The dimension of this subspace is equal to $m - k$. But in every $(m - k)$-dimensional subspace, there are

$$T_n = N^{(m-k)k-k^2}(1 + O(1/N))$$

k-dimensional subspaces. The complexity $K(y_n \,|\, x_n)$ equals the logarithm of T_n with accuracy up to $O(\log n)$. Hence,

$$I(x_n : y_n) = K(y_n) - K(y_n \,|\, x_n) = \log\left(\frac{Q_n}{T_n}\right) + O(\log n).$$

By the direct calculation of the logarithms of Q_n and T_n, we obtain the required statement. $\triangle$

Let us assume that m, k, and F_n ($|F_n| = 2^{\Theta(n)}$) are chosen in such a way that

$$K(x_n) = n + O(\log n), \tag{10}$$

$$K(y_n) = n + O(\log n), \tag{11}$$

$$I(x_n : y_n) = an + O(\log n), \tag{12}$$

where a is some positive constant. Thus, the mutual information of x_n and y_n grows linearly in n. One can easily note that as the ratio k/m tends to $1/2$, the corresponding value of a tends to unity. Consequently, we can choose the values of the parameters m and k in such a way that a will be made arbitrarily close to unity.

Let us show that for the sequences x_n, y_n constructed, their mutual information cannot be materialized; i.e., the assertion of Theorem 2 is valid for them.

Proposition 4. *For the sequences $\boldsymbol{x}$, $\boldsymbol{y}$ that we have constructed and for any sequence $\boldsymbol{z}$ simple conditioned on $\boldsymbol{x}$ and on $\boldsymbol{y}$, the equality $K(z_n) = O(\log n)$ holds.*

Proof. Let z_n be a sequence of words simple conditioned on x_n and y_n (that is, $K(z_n \mid x_n) = O(\log n)$ and $K(z_n \mid y_n) = O(\log n)$). Let us prove that $K(z_n) = O(\log n)$. Fix a positive integer n. In what follows, for the sake of simplicity of notations, we omit the subscript n each time when it does not lead to confusion.

Since $\boldsymbol{z}$ is simple conditioned on $\boldsymbol{x}$ and $\boldsymbol{y}$,

$$\begin{aligned} K(x \mid z) &= K(x,z) - K(z) + O(\log n) \\ &= K(x) + K(z \mid x) - K(z) + O(\log n) = K(x) - K(z) + O(\log n). \end{aligned} \tag{13}$$

A similar calculation can also be performed for y. Put

$$D = \max\{K(x \mid z), K(y \mid z)\}.$$

(Note that $|K(x \mid z) - K(y \mid z)| = O(\log n)$.) In the new notation,

$$K(x \mid z) \le D, \qquad K(y \mid z) \le D.$$

Below we will show that there exist sufficient enough words whose complexity conditioned on z does not exceed D (and thus the number D is large enough). Moreover, we will show that $D = K(x) - O(\log n)$. This means that the conditional complexity $K(x \mid z)$ differs very little from the unconditional complexity $K(x)$. Next, using (13), we will obtain a logarithmic estimate for the complexity of z.

In this proof, we consider chains of subspaces, i.e., finite sequences

$$x^0 - y^1 - x^1 - y^2 - \ldots - y^r - x^r, \tag{14}$$

where x^i, y^i are k-dimensional subspaces of V_n and any adjacent subspaces in the chain are orthogonal. The subspace x_0 is called the left end, and the subspace x^r, the right end of the chain. The number r is called the length of the chain (r is independent of n). We are only interested in chains such that $x^0 = x$. Such a chain is a trajectory of a random walk along the graph G_n. Note that the number of steps in the walk is even, odd steps are marked by x^i, $i = 0, 1, \ldots, r$, and even by y^i, $i = 1, \ldots, r$, respectively.

A random walk along the graph corresponds to the uniform distribution on the set of chains with a fixed left end. The uniform distribution on the chains induces some distribution on the set of their right ends. We will select a value of the parameter r such that the resulting distribution on the set of right ends of the chains will appear to be close to the uniform one. At the same time, we will show that, with a large enough probability, the right end of a randomly chosen chain has a complexity not greater than D conditioned on z. We will thereby show that the number of right ends of chains with complexity not greater than D conditioned on z is large.

First of all, let us show that for a polynomial part of all chains of the form (14), the complexity of the right end x^r is small conditioned on z.

Lemma 5. *Let X^r be the set of all chains of the form* (14). *Then the number of chains whose right ends satisfy the inequality*

$$K(x^r \mid z) \le D$$

is not less that $\dfrac{|X^r|}{\mathrm{poly}(n)}$, where $\mathrm{poly}(n)$ is a polynomial.

Proof. Let us prove a stronger statement. Namely, let us show that not less than a polynomial part consists of chains of subspaces which satisfy the following two conditions:

(a) *all* elements of the chain x^i, y^i have complexity not greater than D conditioned on z;

(b) each pair of adjacent (in the chain) subspaces $\langle y^j, x^j\rangle$ or $\langle x^j, y^{j+1}\rangle$ is random, i.e., has complexity not less than $\log P_n - O(\log n)$ (here and throughout what follows, we use the notation from the proof of Proposition 3).

Note that a pair $\langle y^j, x^j\rangle$ has complexity close to P_n if and only if the complexity $K(x^j)$ is close to $\log Q_n$ and the conditional complexity $K(y^j \mid x^j)$ is close to $\log T_n$. More exactly, randomness of the pair $\langle y^j, x^j\rangle$ is equivalent to the fact that some constant C satisfies the inequalities

$$K(x^j) \geq \log Q_n - C\log n,$$
$$K(y^j \mid x^j) \geq \log T_n - C\log n.$$

Let us prove the lemma by induction on the length of a chain. Let there be not less than $\dfrac{|X^i|}{n^c}$ chains of length i satisfying the condition of the lemma. Choose any of them and consider all its possible extensions $\ldots - y^{i+1} - x^{i+1}$. In doing so, the subspace y^{i+1} must be orthogonal to x^i, and x^{i+1} to y^{i+1}. In all, there are T_n^2 such extensions. It suffices to show that at least a polynomial part of these extensions satisfies conditions (a) and (b).

By the assumption, $K(x^i \mid z)$ and $K(y^i \mid z)$ do not exceed D and the pair $\langle x^i, y^i\rangle$ is random. (For $i = 0$, we assume that $y^0 = y$.) From the definition of D and the relation (13), we obtain

$$\begin{aligned} K(z \mid x^i) &= K(x^i, z) - K(x^i) + O(\log n) \\ &= K(x^i \mid z) + K(z) - K(x^i) + O(\log n) \\ &\leq D + K(z) - K(x) + O(\log n) = O(\log n). \end{aligned}$$

Consider the set L of all k-dimensional subspaces $\widehat{y}$ orthogonal to x^i and such that

$$K(\widehat{y} \mid z) \leq D.$$

Note that if we know the word x^i, we can obtain z with a logarithmic complexity and then initiate the process of enumerating the set L. But the subspace y^i lies in L. Therefore, to find y^i, it suffices to have a program that enumerates L and to know the index of y^i in this enumeration. Thus,

$$K(y^i \mid x^i) \leq \log|L| + O(\log n).$$

By the induction assumption, the pair $\langle x^i, y^i\rangle$ is random, and the complexity of y^i conditioned on x^i is not less than $\log T_n - C\log n$. Consequently, $|L| \geq T_n/\mathrm{poly}(n)$.

Now let us choose some constant $C' > C$ and discard those subspaces of L whose complexity conditioned on x^i is less than $T_n - C'\log n$. More exactly, let $L' \subset L$ consist of all subspaces $\widehat{y}$ such that

$$K(\widehat{y} \mid x^i) \geq T_n - C'\log n.$$

If the constant C' is large enough, then $|L'| \geq T_n/\mathrm{poly}(n)$. We can take any subspace of L' as y^{i+1}. Indeed, for any $\widehat{y} \in L'$ the pair $\langle x^i, \widehat{y}\rangle$ is random, and $K(\widehat{y} \mid z) \leq D$.

In a similar way, we can prove that if $y^{i+1} \in L'$ is chosen, then there are not less than $T_n/\mathrm{poly}(n)$ subspaces $\widehat{x}$, each of which can be taken as x^{i+1}.

Thus, among T_n^2 extensions of the chosen chain of length i, there are not less than $\dfrac{T_n^2}{\mathrm{poly}(n)}$ of those satisfying conditions (a) and (b). (Note that the degree of the resulting polynomial $\mathrm{poly}(n)$ depends on r.) $\triangle$

Let us define a sequence of numbers l_i by the following recurrent relation:

$$l_0 = k, \tag{15}$$

$$l_{i+1} = \max\{l_i + 2k - m; 0\}. \tag{16}$$

Note that, starting from some number, all the numbers l_i equal zero.

We call a chain of subspaces *regular* if for $i = 1, 2, \ldots, r$

$$\dim(x^0 \cap x^i) = l_i. \tag{17}$$

Now let us show that a randomly chosen chain is regular with probability exponentially close to unity. More exactly, we have the following lemma.

Lemma 6. *Among all chains of subspaces of the form* (14), *the part of regular ones is not less than* $1 - 2^{-cn}$ *for some* $c > 0$.

Proof. First of all, let us prove two simple combinatorial sublemmas.

Sublemma 1. *Let us randomly (with respect to the uniform distribution) choose a system of* q *equations in* s *variables over the field* F_n. *Then the rank of the given system will equal* $\min\{s; q\}$ *with probability not less than* $1 - 2^{-cn}$ *(for some constant* $c > 0$*).*

Proof. Let $q \le s$. Let us show that all equations of the system are linearly independent with probability exponentially close to unity. Indeed, if the equations are linearly dependent, then one of them is a linear combination of the others. In all, there are $|F_n|^{q-1}$ linear combinations of $(q-1)$ equations. A single equation can be chosen in $|F_n|^s$ ways. Thus, for each i, the probability that the ith equation of the system is a linear combination of the others does not exceed $|F_n|^{(q-1)-s}$. It remains to sum up the given probabilities over all i from 1 to q. Since $q \le s$, the part of linearly dependent systems does not exceed $q|F_n|^{-1}$. But $|F_n| = 2^{\Theta(n)}$, which proves the statement.

If $q > s$, then the first s equations are linearly independent and the rank of the system equals s with probability exponentially close to unity. $\triangle$

Remark 4. Note an obvious consequence of the sublemma proved. Let us randomly choose a system of q linear equations in s variables over a finite field F. Assume that we know that some sets of equations from the system obtained are linearly independent. It is clear that under the given condition, the probability of the event "the rank of the whole system chosen equals $\min\{s; q\}$" is all the more exponentially close to unity.

Sublemma 2. (1) *Let* W *be a linear space over a finite field* F, *and let* a *be an* s_1*-dimensional subspace of* W. *Let us randomly (with respect to the uniform distribution) choose an* s_2*-dimensional subspace* b *in* W. *Then the probability of the event that* $\dim(a \cap b) = r$ *depends solely on the values of* r, s_1, s_2, $\dim(W)$, *and* $|F|$ *(but does not depend on the choice of the* s_1*-dimensional space* a*).*

(2) *Let* W *be a linear space over a finite field* F, *and let* a *and* b *be* s*-dimensional linear subspaces of* W. *Let us randomly choose an* s*-dimensional subspace* a_1 *from the orthogonal complement of* a *and an* s*-dimensional subspace* b_1 *from the orthogonal complement of* b. *Then, for any* l, *the probabilities of the events that* $\dim(a_1 \cap b) = l$ *and* $\dim(a \cap b_1) = l$ *are equal.*

Proof. (1) Let a' be an arbitrary linear subspace of W of dimension s_1. Obviously, there exists an automorphism φ of the space W such that $\varphi a = a'$. Then for any linear subspace b of the space W we have $\dim(a \cap b) = \dim(a' \cap \varphi b)$. Consequently,

$$\mathrm{Prob}_b[\dim(a \cap b) = l] = \mathrm{Prob}_b[\dim(a' \cap \varphi b) = l] = \mathrm{Prob}_b[\dim(a' \cap b) = l].$$

(2) Note that $\dim(a^\perp \cap b) = \dim(a \cap b^\perp)$. Indeed,

$$\dim(W) - \dim(a \cap b^\perp) = \dim\left((a \cap b^\perp)^\perp\right) = \dim(a^\perp \oplus b)$$
$$= \dim(a^\perp) + \dim(b) - \dim(a^\perp \cap b) = \dim(W) - \dim(a^\perp \cap b),$$

where $(a^\perp \oplus b)$ denotes the sum of the subspaces $a^\perp$ and b. It remains to apply statement (1) of Sublemma 2 to the space $a^\perp$, which includes $b \cap a^\perp$ and the randomly chosen a_1, and to the space $b^\perp$, which includes $a \cap b^\perp$ and the randomly chosen b_1. $\triangle$

Let us prove the lemma by induction on the chain length. The induction base is obvious. To perform the induction step, it suffices to show that if the chain

$$x^0 - y^1 - x^1 - y^2 - \ldots - y^i - x^i$$

is regular, then all its extensions

$$x^0 - y^1 - x^1 - y^2 - \ldots - y^i - x^i - y^{i+1} - x^{i+1},$$

except for an exponentially small part, are also regular. Consider the final fragment of the given chain,

$$x^i - y^{i+1} - x^{i+1}.$$

To choose y^{i+1} and x^{i+1}, we have T_n^2 different possibilities (in the notation of the proof of Proposition 3). Let us calculate the probability that $\dim(x^0 \cap x^{i+1}) = l_{i+1}$.

Assume that y^{i+1} is already chosen. Now we may consider only the pair of the subspaces x^0 and y^{i+1}: we need to know what is the probability of the event that a randomly chosen third subspace (we call it x^{i+1}) has an l_{i+1}-dimensional intersection with x^0 provided that y^{i+1} and x^{i+1} are orthogonal. By statement (2) of Sublemma 2, the latter problem is equivalent to another one, namely: What is the probability that a randomly chosen subspace y^0 has an l_{i+1}-dimensional intersection with y^{i+1} provided that x^0 and y^0 are orthogonal?

Thus, in order to calculate the probability with which the equality $\dim(x^0 \cap x^{i+1}) = l_{i+1}$ holds, we may solve another problem as follows. Let y^0 be a randomly chosen k-dimensional subspace of V_n orthogonal to x^0, and let y^{i+1} be a randomly chosen k-dimensional subspace of V_n orthogonal to x^i. What is the probability that $\dim(y^0 \cap y^{i+1}) = l_{i+1}$?

The subspaces y^0 and y^{i+1} can be defined by systems of $m - k$ linear equations. Without loss of generality, we may assume that the first k equations in the first system correspond to the orthogonality of y^0 to the space x^0, and the first k equations in the second system correspond to the orthogonality of y^{i+1} to the space x^i. We may also assume that the first l_i equations in both systems are identical and correspond to the requirement of orthogonality of both subspaces y^0 and y^{i+1} to the subspace $x^0 \cap x^i$.

Combining the systems of equations that define y^0 and y^{i+1}, we obtain a system of $2(m-k) - l_i$ equations. By Sublemma 1, with probability exponentially close to unity, the given system has rank $\min\{m, 2(m-k) - l_i\}$ and, hence,

$$\dim(y^0 \cap y^{i+1}) = \max\{0; 2k - (m - l_i)\} = l_{i+1}.$$

Therefore, with the same probability we have $\dim(x^0 \cap x^{i+1}) = l_{i+1}$. Thus, with probability exponentially close to unity, the chain satisfies the condition (17). $\triangle$

Let A^r be the set of all k-dimensional subspaces of V_n whose intersection with x has dimension l_r. By Lemma 6, for most chains of the form (14), x^r belongs to A^r. Now let us prove the following property of homogeneity: all subspaces from A^r are right ends of the same number of regular chains of length r (as before, we consider only chains whose left end coincides with x).

Lemma 7. *If $v, w \in A^r$, then the number of regular chains whose right end coincides with v equals the number of regular chains whose right end coincides with w.*

Proof. Let us prove this by induction on r. For each subspace $x^r \in A^r$, we need to calculate the number of chains

$$x^{r-1} - y^r - x^r,$$

where $x^{r-1} \in A^{r-1}$. More exactly, it suffices to show that the number of such chains does not depend on the choice of $x^r \in A^r$ and depends on the index r only. Let us fix the subspace $x^r \in A^r$ and choose a random y^r orthogonal to it. Next, let us choose a random x^{r-1} orthogonal to y^r. We want to know what is the probability of the event that $x^{r-1} \in A^{r-1}$ (i.e., $\dim(x^0 \cap x^{r-1}) = l_{r-1}$).

Let the subspace y^r be already chosen. We need to determine the probability for a randomly chosen subspace x^{r-1} to have an l_{r-1}-dimensional intersection with x^0 provided that x^{r-1} and y^r are orthogonal. By statement (2) of Sublemma 2, the probability in question is equal to that of the event that a randomly chosen subspace y^0 has an l_{r-1}-dimensional intersection with y^r provided that y^0 and x^0 are orthogonal.

Thus, we may proceed to the solution of a new problem, namely, to find the probability that a pair of randomly chosen k-dimensional subspaces y^0 and y^r, the first of which is orthogonal to x^0 and the second is orthogonal to x^r, have an l_{r-1}-dimensional intersection.

The subspaces y^0 and y^r are defined by systems of $n-k$ linearly independent equations. We assume that the first k equations in the first system correspond to the orthogonality of y^0 to the space x^0, and the first k equations in the second system correspond to the orthogonality of y^r to the space x^r. Moreover, we may assume that the first l_r equations of both systems are identical (and corresponding to the orthogonality of both subspaces y^0, y^r to the subspace $x^0 \cap x^r$). Let us combine the two given systems of equations. The space of solutions of the new system (of $2(n-k) - l_r$ equations) is the intersection of y^0 and y^r. We are interested in the probability that its dimension equals l_{r-1}. Obviously, this probability is determined by the values of n, k, and r, but is independent of the choice of a specific x^r. We do not need to know the magnitude of the given probability. It counts only that it is uniquely determined by the index r (for a given n). $\triangle$

Now we complete the proof of Proposition 4. Consider the set X^r of subspace chains (14) with length r. According to Lemma 6, all such chains, except for an exponentially small part, are regular, which means that their right ends lie in A^r. Furthermore, by Lemma 5, at least $\dfrac{|X^r|}{\mathrm{poly}(n)}$ of these chains have right ends simple conditioned on z (i.e., their complexity conditioned on z does not exceed D). Therefore, not less than $\dfrac{|X^r|}{\mathrm{poly}(n)}$ chains both are regular and have right ends of complexity not greater than D conditioned on z. But because of the homogeneity (Lemma 7), all subspaces from A^r are right ends of the same number of regular chains. Consequently, not less than $\dfrac{|A^r|}{\mathrm{poly}(n)}$ elements of the set A^r have complexity not greater than D conditioned on z.

Let us choose the minimum r for which $l_r = 0$. Then A^r consists of all k-dimensional subspaces having zero intersection with x. In this case, all k-dimensional subspaces of V_n, except for an exponentially small part, lie in A^r. Indeed, a subspace x of V_n is determined by a system of $(m-k)$ linear equations. Let us randomly choose additional $(m-k)$ linearly independent equations (which define some k-dimensional subspace of V_n). Let us combine the two given systems of equations. The new system contains $2(m-k)$ equations. Since $2k < m$, with probability exponentially close to unity, the rank of this system equals m (Sublemma 1), and it has no nontrivial solutions. This means that a randomly chosen subspace of V_n has a zero intersection with x with probability exponentially close to unity.

Thus, A^r contains all k-dimensional subspaces of V_n except for an exponentially small part. We know that not less than a polynomial part of all subspaces from A^r have complexity not greater than D_n conditioned on z. Therefore, at least a polynomial part among all k-dimensional subspaces

of V_n, as well, have complexity not greater than D conditioned on z. Thus, if Q_n is the number of all k-dimensional subspaces of V_n, we have

$$2^D \geq \frac{Q_n}{\mathrm{poly}(n)}.$$

Let us recall that $K(x) = \log Q_n + O(\log n)$. Moreover, $D = K(x \mid z) + O(\log n) = K(x) - K(z) + O(\log n)$. Hence,

$$K(x) - K(z) \geq K(x) - O(\log n).$$

Thus, $K(z) = O(\log n)$. $\triangle$

Remark 5. The construction considered may be generalized. Let k_1, k_2, m be positive integers such that $k_1 + k_2 < m$. Consider word sequences $\{x_n\}$ and $\{y_n\}$, where x_n and y_n are random orthogonal linear subspaces of F_n^m, whose dimensions are equal to k_1 and k_2 respectively.

Then for some a, b, c (which are determined by the parameters k_1, k_2, and m),

$$K(x_n) = bn + O(\log n), \qquad K(y_n) = cn + O(\log n), \qquad I(x_n : y_n) = an + O(\log n).$$

The consideration analogous to the proof of Proposition 4 shows that the mutual information between $\{x_n\}$ and $\{y_n\}$ cannot be materialized, i.e., for any sequence $\{z_n\}$ simple conditioned on $\{x_n\}$ and $\{y_n\}$, we have $K(z_n) = O(\log n)$.

The author thanks N. K. Vereshchagin for the scientific supervision and help in the work on the paper.

REFERENCES

1. Gács, P. and Körner, J., Common Information Is Far Less than Mutual Information, *Probl. Control Inf. Theory*, 1973, vol. 2, no. 2, pp. 149–162.
2. Kolmogorov, A.N., Combinatorial Foundations of Information Theory and Probability Calculus, *Usp. Mat. Nauk*, 1983, vol. 38, no. 4, pp. 27–36.
3. Zvonkin, A.K. and Levin, L.A., Complexity of Finite Objects and Foundations of the Notions of Information and Randomness Using the Theory of Algorithms, *Usp. Mat. Nauk*, 1970, vol. 25, no. 6, pp. 85–127.
4. Muchnik, An.A., On Separation of Common Information of Two Words, *Pervyi Vsemirnyi kongress obshchestva matematicheskoi statistiki i teorii informatsii im. Bernoulli, Tez. Dokl.* (Proc. First World Congress Bernoulli Society Math. Stat. Prob. Theory), Moscow, 1986, p. 453.
5. Muchnik, An.A., On Common Information, *Theor. Comput. Sci.*, 1998, vol. 207, pp. 319–328.
6. Muchnik, An.A., Shen, A., Romashchenko, A., and Vereshchagin, N.K., Upper Semi-Lattice of Binary Strings with the Relation x is Simple Conditional to y, *Preprint of DIMACS TR 97-74, Rutgers Univ.*, Piscataway, NJ, 1997.
7. Hammer, D., Romashchenko, A., Shen, A., and Vereshchagin, N., Inequalities for Shannon Entropies and Kolmogorov Complexities, *Proc. Twelfth Annual IEEE Conf. on Computational Complexity*, Ulm, 1997, pp. 13–23.