On joint conditional complexity (entropy).

Andrey A. Muchnik* Nikolay K. Vereshchagin†

Abstract

Conditional Kolmogorov complexity of a word a given a word b is the minimum length of a program that prints a given b as an input. We generalize this notion to quadruples of strings a, b, c, d: their joint conditional complexity $K((a \to c) \wedge (b \to d))$ is defined as the minimum length of a program that given a outputs c and given b outputs d. In this paper, we prove that the joint conditional complexity cannot be expressed in terms of usual conditional (and unconditional) Kolmogorov complexity. This result provides a negative answer to the following question, asked by A.Shen on a session of Kolmogorov seminar at Moscow State University in 1994: Is there a problem of information processing whose complexity is not expressible in terms of conditional (and unconditional) Kolmogorov complexity?

We show that a similar result holds for classical Shannon entropy. We provide two proofs of both results, an effective one and a "quasi-effective" one. Finally we present a quasi-effective proof of a strong version of the following statement: there are two strings whose mutual information cannot be extracted. Previously, only a non-effective proof of that statement was known.

The results concerning Kolmogorov complexity appeared, in a preliminary form, in the Proceedings of the 16th Annual IEEE Conference on Computational Complexity in 2001. [A. Muchnik and N. Vereshchagin. "Logical operations and Kolmogorov complexity. II". Proc. of 16th Annual IEEE Conference on Computational Complexity, Chicago, June 2001, pp. 256–265.]

1 Introduction

Let $V(x, y)$ be a partial computable function mapping pairs of binary strings to binary strings. *The conditional Kolmogorov complexity* $K_V(b|a)$ is defined as the minimal length of a string p with $V(p, a) = b$. If this equality holds, we say that p is *a description of b when a is known, with respect to the conditional description*

*Institute of New Technologies in Education, 10 Nizhnyaya Radischewskaya, Moscow, Russia 109004

†Dept. of Mathematical Logic and Theory of Algorithms, Moscow State University, Leninskie Gory, Moscow 119991, Russia; e-mail: ver@mccme.ru. The author was partially supported by Russian Foundation for Basic Research grant 09-01-00709.

mode V. By Solomonoff–Kolmogorov theorem there is a description mode U that is asymptotically not worse than any other description mode V. This means that that for all partial computable function V there is a number c such that $K_U(b|a) \leqslant K_V(b|a) + c$ for all a, b. Any function U that satisfies Solomonoff–Kolmogorov theorem is called *optimal*. We fix an optimal description mode U and use the notation $K(b|a)$ for $K_U(b|a)$ dropping the subscript U. If we replace U by another optimal description mode then $K(b|a)$ may change by at most a bounded term.

Kolmogorov complexity $K(a)$ is defined as $K(a|\text{empty word})$. If $V(p, \text{empty word}) = a$, we say that p is *a description of a with respect to description mode V*. Kolmogorov complexity of every string is finite. Moreover, it exceeds the length of the string by at most a constant. Those strings whose complexity is close to the length are called random.

Kolmogorov complexity cannot increase via computable transformations: for every partial computable function f there is a constant c such that

$$K(f(x)) \leqslant K(x) + c$$

for all x in the domain of f.

Kolmogorov complexity of the pair $\langle p, q\rangle$ of strings is defined as as the complexity of the string $[p, q]$, where $\langle x, y\rangle \mapsto [x, y]$ stands for a fixed computable bijection between the set of pairs of binary strings and the set of binary strings. If we replace the fixed bijection by another one, $K([p, q])$ will change by at most a constant, as computable transformations can increase complexity only by a constant. Therefore the choice of computable bijection does not matter. We will abbreviate $K(\langle p, q\rangle) = K([p, q])$ by $K(p, q)$.

One may expect that the complexity of the pair $\langle p, q\rangle$ does not exceed the sum of complexities of p and q (up to an additive constant): $K(p, q) \leqslant K(p) + K(q)$. However this inequality holds only up to a logarithmic term: $K(p, q) \leqslant K(p) + K(q) + O(\log K(q))$. The extra term is necessary because we cannot just concatenate the minimal length descriptions of p and q to get a description of the pair $\langle p, q\rangle$. We need also know the length of one of these descriptions (to parse the concatenation into two descriptions). This extra information can be encoded in a "self-delimiting" way in $O(\log K(q))$ (or $O(\log K(p))$) bits (this means that reading the encoding we can find the place where it ends). Then we can prefix the concatenation of descriptions of p and q with the encoding of the length of one of them.

A *programming language* is a partial computable function P from $\{0,1\}^* \times \{0,1\}^*$ to $\{0,1\}^*$ that has the following property. For any other partial computable function $Q(x, y)$ there exists a total computable function s such that $P(s(x), y) = Q(x, y)$ for all x, y. (The function s is called a translator from Q-programs to P-programs.) If $P(x, y) = z$ then we say that *the program x outputs z given y as input*, or just that *the program x transforms (or maps) y into z*. We fix any programming language P and we denote by $a \to b$ the set of all programs that transform a into b. We will also use the abbreviation $[x](y)$ for $P(x, y)$.

Kolmogorov complexity of a non-empty set A of strings is defined as minimal complexity of its elements: $K(A) = \min\{K(a) \mid a \in A\}$. If A is a set of all programs that solve certain algorithmic problem, we will identify A with that problem and call elements of A its solutions. For example, the set $a \to b$ is identified with the algorithmic problem of transforming a into b and $K(a \to b)$ is the complexity of that problem.

Recall that Kolmogorov complexity does not increase via computable transformations. Therefore, if we replace the fixed programming language P by another programming language Q, the quantity $K(a \to b)$ cannot change by more than an additive constant not depending on a, b. This can be shown also in the following way. It is easy to see that $K(a \to b)$ is equal to $K(b|a)$ up to an additive constant. As this equality holds for every programming language P used in the definition of the set $a \to b$, the quantity $K(a \to b)$ does not depend on the choice of P.

The complexity of the problem of joint transformation of a to c and b to d can be defined in two equivalent ways:

- $K((a \to c) \wedge (b \to d))$ is the minimal complexity of a program that transforms $0a$ to c and transforms $1b$ to d, that is, the complexity of intersection of the sets $0a \to c$ and $1b \to d$;
- $K((a \to c) \wedge (b \to d))$ is the minimal complexity of the pair $\langle p, q\rangle$, where p is a program transforming a to c, and q is a program transforming b to d, that is, the complexity of the Cartesian product of the sets $a \to c$ and $b \to d$.

All the above notions have counterparts in the classical information theory. The analog of Kolmogorov complexity of a string is Shannon entropy of a random variable, which is defined as follows. Let α be a random variable and let $a_1, \ldots, a_n$ be its possible outcomes. Let p_i stand for the probability that $\alpha = a_i$. *The Shannon entropy of* α is defined by the formula

$$H(\alpha) = -\sum_i p_i \log p_i$$

(All the logarithms in the paper are base 2 and $0 \log 0 = 0$.) The convexity of the logarithmic function implies that Shannon entropy does not exceed the logarithm of the number of possible outcomes and attains this value if and only if α has the uniform distribution. Shannon entropy is non-negative and equals zero if and only if all probabilities $p_1, \ldots, p_n$, except one, are equal to 0.

Let β be another random variable having $b_1, \ldots, b_k$ as possible outcomes and assume that α and β have a joint distribution. This means that for every pair i, j the probability of the event $\alpha = a_i, \beta = b_j$ is given. For every fixed j the value $H(\alpha|\beta = b_j)$ is defined similar to $H(\alpha)$, but this time, in the formula for entropy, we use conditional probabilities $\Pr[\alpha = a_i|\beta = b_j]$ in place of p_i. *The conditional Shannon entropy* is defined by the expression

$$H(\alpha|\beta) = \sum_j \Pr[\beta = b_j] \cdot H(\alpha|\beta = b_j).$$

Conditional Shannon entropy does not exceed unconditional one: $H(\alpha|\beta) \leqslant H(\alpha)$. Similar inequality holds for Kolmogorov complexity: $K(a|b) \leqslant K(a) + O(1)$.

The definition of conditional Shannon entropy implies that

$$H(\alpha, \beta) = H(\alpha|\beta) + H(\beta),$$

where in the left hand side $H(\alpha, \beta)$ stands for Shannon entropy of the random variable $\langle \alpha, \beta \rangle$. Similar inequality for Kolmogorov complexity

$$K(a, b) = K(a|b) + K(b),$$

holds up to an additive term $O(\log K(a, b))$ and was established by Kolmogorov and Levin [7]. This equality and the inequality $H(\alpha|\beta) \leqslant H(\alpha)$ imply that

$$H(\alpha, \beta) \leqslant H(\alpha) + H(\beta).$$

We have already mentioned similar inequality for Kolmogorov complexity.

Conditional entropy is zero only when all the terms $H(\alpha|\beta = b_j)$ with positive $\Pr[\beta = b_j]$ are equal to zero. In other words, $H(\alpha|\beta) = 0$ if and only if there is a function g such that $\alpha = g(\beta)$ with probability 1.

Recall that $K(a|b)$ is equal (up to a constant) to the minimal complexity of a program that transforms b to a. A similar statement holds for Shannon entropy up to a logarithmic additive term: $H(\alpha|\beta)$ is equal to the minimal entropy of a random variable γ (jointly distributed with α, β) such that for some function g we have $\alpha = g(\beta, \gamma)$ with probability 1. Indeed, for all triples of random variables we have

$$H(\alpha|\beta) \leqslant H(\alpha|\beta, \gamma) + H(\gamma).$$

Therefore the entropy of every random variable γ with $H(\alpha|\beta, \gamma) = 0$ is at least $H(\alpha|\beta)$. The last inequality follows from the equality

$$H(\alpha, \gamma|\beta) = H(\alpha|\beta, \gamma) + H(\gamma|\beta),$$

which can be considered as a conditional version of the equality $H(\alpha, \gamma) = H(\alpha|\gamma) + H(\gamma)$. It can be derived from the latter equality by averaging over all outcomes of β. (Here $H(\alpha, \gamma|\beta)$ stands for $H(\langle \alpha, \gamma \rangle|\beta)$ and $H(\alpha|\beta, \gamma)$ means $H(\alpha|\langle \beta, \gamma \rangle)$.)

The converse inequality follows from the following

Theorem 1 ([5]). *For all jointly distributed α, β there is a random variable γ jointly distributed with α, β such that $H(\gamma) \leqslant H(\alpha|\beta) + O(\log H(\alpha|\beta))$ and $H(\alpha|\beta, \gamma) = 0$.*

Similar to the notion of $K((a \to c) \wedge (b \to d))$, define

$$H((\alpha \to \gamma) \wedge (\beta \to \delta)) = \min\{H(\rho) \mid H(\gamma|\alpha, \rho) = H(\delta|\beta, \rho) = 0\}.$$

Here α, γ, β and δ are given jointly distributed random variables, and ρ ranges over all random variables jointly distributed with the given variables. In other

words, we minimise over all probability distributions of 5-tuples $\langle \alpha, \gamma, \beta, \delta, \rho \rangle$ whose projections on the first 4 coordinates yield the given distribution.

The following natural question arises: are the values $K((a \to c) \wedge (b \to d))$ and $H((\alpha \to \gamma) \wedge (\beta \to \delta))$ essentially novel? That is, can they be expressed in terms of complexities of a, b, c, d (entropies of $\alpha, \gamma, \beta, \delta$) and mutual conditional complexities (entropies) of tuples composed of them? To formulate both question more rigorously, let us define *the complexity vector* of a tuple of strings $\langle a_1, \dots, a_k \rangle$ as the vector of length $2^k - 1$ consisting of complexities of $a_1, \dots, a_k$, their pairs, triples etc. For instance, the complexity vector of the triple $\langle a, b, c \rangle$ is defined as $\langle K(a), K(b), K(c), K(a,b), K(a,c), K(b,c), K(a,b,c) \rangle$. If we replace in this definition binary strings by random variables and Kolmogorov complexity by Shannon entropy, we obtain a similar definition of *the entropy vector* of a tuple of jointly distributed random variables $\langle \alpha_1, \dots, \alpha_k \rangle$. Note that complexity (entropy) vector determines also all the mutual conditional complexities (entropies) of tuples composed of given strings (random variables). For Kolmogorov complexity this holds with logarithmic precision by Kolmogorov–Levin equality $K(\langle a, b \rangle) = K(a|b) + K(b) + O(\log K(\langle a, b \rangle))$. For Shannon entropy this is implied by the equality $H(\alpha, \beta) = H(\beta) + H(\alpha|\beta)$.

The answer to this question is provided by the following theorem.

Theorem 2. *There are two sequences of quadruples of words $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ and $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$ whose complexities are linear in n and such that the following is true. The corresponding components of the complexity vector of $\tilde{a}_n, \tilde{b}_n, \tilde{c}_n, \tilde{d}_n$ and $\bar{a}_n, \bar{b}_n, \bar{c}_n, \bar{d}_n$ differ by at most $O(1)$, while the difference $K((\tilde{a}_n \to \tilde{c}_n) \wedge (\tilde{b}_n \to \tilde{d}_n)) - K((\bar{a}_n \to \bar{c}_n) \wedge (\bar{b}_n \to \bar{d}_n))$ grows linearly in n. The similar statement holds for Shannon entropy.*

This theorem answers in negative the following question asked by A. Shen on a session of Kolmogorov seminar at Moscow State University in 1994: is it true that Kolmogorov complexity of every set obtained from singletons $\{a_1\}, \dots, \{a_k\}$ using operations $\vee, \wedge, \to$ is determined by the complexity vector of $\langle a_1, \dots, a_k \rangle$ (with logarithmic precision)? The set operations $\vee, \wedge, \to$ are defined as follows:

- $A \wedge B = \{\langle a, b \rangle \mid a \in A, b \in B\}$,
- $A \vee B = \{0a \mid a \in A\} \cup \{1b \mid b \in B\}$,
- $A \to B = \{p \mid [p](x) \in B \text{ for every } x \in A\}$.

For all expressions having been studied by 1994 the answer to Shen's question is positive. Here is the list of those of them for which the answer is not trivial. All the equalities and inequalities below hold up to an additive term $O(\log K(a, b, \dots))$. (To simplify notation we write a in place of $\{a\}$ to denote the singleton set.)

- $K((a \to b) \wedge c) = K(c) + K(b|a, c)$.
  The upper bound $K((a \to b) \wedge c) \leqslant K(c) + K(b|a, c)$ is trivial. The

lower bound is shown as follows. Let p be a program transforming a to b for which $K(p,c)$ is minimal. We need to prove that $K(p,c) \geqslant K(c) + K(b|a,c)$. Obviously, $K(b|a,c) \leqslant K(p|c)$ (any program q mapping c to p can be transformed to a program mapping $\langle a,c\rangle$ to b: apply q to c to find p, then apply p to a to find b). It remains to use the equality $K(p|c) = K(p,c) - K(c)$.

- $K((a \to b) \wedge (b \to a)) = \max\{K(b|a), K(a|b)\}$.
  The lower bound $K((a \to b) \wedge (b \to a)) \geqslant \max\{K(b|a), K(a|b)\}$ is obvious; the upper bound was proven in [1].

- $K((a \to c) \wedge (b \to c)) = K((a \vee b) \to c) = \max\{K(c|a), K(c|b)\}$.
  Here the lower bound $K((a \to c) \wedge (b \to c)) \geqslant \max\{K(c|a), K(c|b)\}$ is also obvious and the upper bound was established in [4].

- $K((a \to b) \wedge (b \to c)) = \max\{K(b,c|a), K(c|b)\}$.
  Any solution $\langle p,q\rangle$ to this problem may be transformed into an element of the set $a \to (b \wedge c)$, as $c = [q]([p](a))$. Therefore this problem is equivalent to the problem $(a \to (b \wedge c)) \wedge (b \to c)$. (Two sets are called equivalent if there is an algorithm that for all a, b, c given any element of the first set computes an element of the second set, and vice versa. Obviously, complexities of equivalent sets differ by at most $O(1)$.) The latter problem is equivalent to the problem $(a \to (b \wedge c)) \wedge (b \to (b \wedge c))$. Using the previous item we see that its complexity is equal to $\max\{K(b,c|a), K(b,c|b)\}$.

- $K((a \to b) \to c) = \min\{K(c), K(a) + K(c|a,b)\}$.
  The upper bound is obvious and the lower bound can be shown using the method from the paper [6] (where a weaker inequality $K((a \to b) \to c) \geqslant \min\{K(c), K(a)\}$ was proved).

2 The complexity of the problem $(a \to c) \wedge (b \to d)$

The best upper and lower bounds of the complexity of the problem $(a \to c) \wedge (b \to d)$ known to the authors are presented in the following theorem.

Theorem 3.

$$K((a \to c) \wedge (b \to d)) \leqslant \min\{K(c|a) + K(d|b), K(d|b,c) + K(c), K(c|a,d) + K(d)\},$$
$$K((a \to c) \wedge (b \to d)) \geqslant \max\{K(b,c,d|a) - K(b|a,c), K(a,c,d|b) - K(a|b,d)\}.$$

Proof. By definition a solution to this problem is a pair $\langle p,q\rangle$ of programs such that p maps a to c, and q maps b to d. Taking as p and q the shortest such programs we get the upper bound $K(c|a) + K(d|b)$ for its complexity.

To prove the upper bound $K(c) + K(d|b,c)$ let p be the program mapping all inputs to c and let q be the program pairing its input with c and applying to the resulting pair a minimum length program mapping $\langle b,c\rangle$ to d. The upper bound $K(d) + K(c|a,d)$ is proven in a similar way.

To prove the lower bound let $\langle p, q\rangle$ be any element of the set $(a \to c) \wedge (b \to d)$. Consider the triple $\langle p, q, r\rangle$, where r is a shortest program mapping $\langle a, c\rangle$ to b. Given this triple and a, we can find $\langle b, c, d\rangle$: apply p to a to obtain c, then apply r to $\langle a, c\rangle$ to get b, finally apply q to b to obtain d. Therefore

$$K(b, c, d|a) \leqslant K(p, q) + K(r) = K(p, q) + K(b|a, c),$$

hence $K(b, c, d|a) - K(b|a, c) \leqslant K(p, q)$. The bound $K(a, c, d|b) - K(a|b, d) \leqslant K(p, q)$ is proven in a similar way. □

A theorem similar to Theorem 3 holds for Shannon entropy and is proved by similar arguments. To prove the first inequality

$$H((\alpha \to \gamma) \wedge (\beta \to \delta)) \leqslant H(\gamma|\alpha) + H(\delta|\beta) + O(\log H(\alpha, \beta, \gamma, \delta)) \qquad (1)$$

let ρ, σ be random variables of minimal Shannon entropy such that $H(\gamma|\alpha, \rho) = H(\delta|\beta, \sigma) = 0$. By Theorem 2 their entropies are close to $H(\gamma|\alpha)$, $H(\delta|\beta)$, respectively. More specifically, by this theorem we obtain jointly distributed random variables γ, α, ρ and jointly distributed random variables δ, β, σ such that the above inequalities hold. It is easy to see that these distributions can be extended to a joint distribution of $\alpha, \beta, \gamma, \delta, \rho, \sigma$ such that the joint distribution of the initial four random variables coincides with the given one. Then the random variable $\langle \rho, \sigma\rangle$ witnesses the inequality (1).

The second inequality

$$H((\alpha \to \gamma) \wedge (\beta \to \delta)) \leqslant H(\delta|\beta, \gamma) + H(\gamma) + O(\log H(\alpha, \beta, \gamma, \delta)),$$

in the upper bound, as well as the third one, is proven in a similar way.

The first inequality in the lower bound

$$H((\alpha \to \gamma) \wedge (\beta \to \delta)) + H(\beta|\alpha, \gamma) \geqslant H(\beta, \gamma, \delta|\alpha)$$

follows from the inequality

$$H(\rho) + H(\beta|\alpha, \gamma) + H(\gamma|\rho, \alpha) + H(\delta|\rho, \beta) \geqslant H(\beta, \gamma, \delta|\alpha),$$

which holds for all random variables and is easy to show. The second inequality in the lower bound is proved by symmetrical arguments.

Example 1. Let a, b, c, d be obtained by cutting a random binary string of length $4n$ in four blocks, each of length n. Then both the lower and upper bounds in the above theorem are equal to $2n$, hence the joint conditional complexity is $2n$. For Shannon entropy, a similar example is a quadruple of independent random variables $\alpha, \beta, \gamma, \delta$, each of them is uniformly distributed among binary strings of length n.

Example 2. It is easy to find strings a, b, c, d for which the lower bound in Theorem 3 is less then the upper bound and the joint conditional complexity is equal to the lower bound. Consider a random string of length $2n$, cut it in two blocks x, y of length n and let $a = d = x$, $b = c = y$. For these

a, b, c, d the lower bound is equal to n, and the upper bound to $2n$. The joint conditional complexity is equal to n, as we can take $p = q = x \oplus y$ (the sign $\oplus$ refers to bitwise addition modulo 2). For Shannon entropy a similar example is provided by independent random variables α, β uniformly distributed among binary strings of length n and by $\gamma = \beta$, $\delta = \alpha$.

2.1 The proof of Theorem 2

Let us call quadruples built in the way used in the above two examples "standard". More specifically, a quadruple $\langle a, b, c, d\rangle$ is standard if a, b, c, d are built from the constant number of blocks of the same random string using concatenation and $\oplus$. We can show that for every standard quadruple a, b, c, d the joint conditional complexity $K((a \to c) \wedge (b \to d))$ is equal to the lower bound of Theorem 3 (up to an additive constant that depends on the number of blocks). Therefore at least one of the quadruples in Theorem 2 must be non-standard and $K((a \to c) \wedge (b \to d))$ must be greater than the lower bound of Theorem 3. In our proof, that quadruple will be $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$. To construct it we will use linear algebra over finite fields in a way similar to that used in [3]. The other quadruple $\bar{a}, \bar{b}, \bar{c}, \bar{d}$ will be standard.

We begin with constructing the standard quadruple $\langle \bar{a}, \bar{b}, \bar{c}, \bar{d}\rangle$. Pick a random binary string of length $7n$ and cut it into 7 blocks u, v, w, p, q, r, s, each of length n. Let $\bar{a} = uvws$, $\bar{b} = pqrs$, $\bar{c} = ups$, $\bar{d} = vqs$. It is easy to see that the complexity vector of the resulting quadruple is the following:

$$K(\bar{a}) = K(\bar{b}) = 4n, \quad K(\bar{c}) = K(\bar{d}) = 3n,$$
$$K(\bar{a}, \bar{b}) = 7n, \quad K(\bar{a}, \bar{c}) = K(\bar{a}, \bar{d}) = K(\bar{b}, \bar{c}) = K(\bar{b}, \bar{d}) = K(\bar{c}, \bar{d}) = 5n,$$
$$K(\bar{a}, \bar{c}, \bar{d}) = K(\bar{b}, \bar{c}, \bar{d}) = 6n, \quad K(\bar{a}, \bar{b}, \bar{c}) = K(\bar{a}, \bar{b}, \bar{d}) = K(\bar{a}, \bar{b}, \bar{c}, \bar{d}) = 7n.$$

This implies that the lower bound in Theorem 3 is $K((\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d})) \geqslant n + O(\log n)$. On the other hand, $K((\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d})) \leqslant n + O(1)$, as given $p \oplus v$ one can map $\bar{a}$ to $\bar{c}$, and $\bar{b}$ to $\bar{d}$.

Now we construct the non-standard quadruple $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$. Let F_n denote the field of cardinality 2^n. We will consider points, lines and planes in the three-dimensional affine space over F_n. There are 2^{3n} points, $2^{4n+o(1)}$ lines (the exact number of lines is $\frac{2^{3n}(2^{3n}-1)}{2^n(2^n-1)}$) and $2^{3n+o(1)}$ planes in this space. As $\langle \tilde{a}, \tilde{b}\rangle$ we take any random pair of (different) intersecting lines, $\tilde{c}$ will be its common point and $\tilde{d}$ its common plane. Let us specify what we mean by a random pair of intersecting lines. There are $2^{7n+o(1)}$ pairs of intersecting lines. Choose an integer constant ε such that $2^{7n-\varepsilon}$ does not exceed this number and call a pair of intersecting lines *random* if its conditional Kolmogorov complexity when n is known is at least $7n - \varepsilon$. Such pair does exist, as the number of descriptions of length less than $7n - \varepsilon$ equals $1 + 2 + \cdots + 2^{7n-\varepsilon-1} < 2^{7n-\varepsilon}$.

It is easy to verify that the complexity vector of $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$ is the same as that of $\bar{a}, \bar{b}, \bar{c}, \bar{d}$ (up to an additive constant). Thus it suffices to show that $K((\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})) > cn$ for some $c > 1$ and all sufficiently large n. (Recall that $K((\bar{a} \to \bar{c}) \wedge (\bar{b} \to \bar{d})) \leqslant n + O(1)$.)

For Shannon entropy, the example is very similar: $\langle\tilde{\alpha},\tilde{\beta}\rangle$ is uniformly distributed among all pairs of (different) intersecting lines, $\tilde{\gamma}$ is its common point, and $\tilde{\delta}$ is the plane that contains both lines. Thus it remains to prove the following theorem.

Theorem 4. *$K((\tilde{a}\to\tilde{c})\wedge(\tilde{b}\to\tilde{d}))\geqslant 1.5n-O(\log n)$ and $H((\tilde{\alpha}\to\tilde{\gamma})\wedge(\tilde{\beta}\to\tilde{\delta}))\geqslant 1.5n-O(1)$.*

Proof. Let us prove the first inequality. Let p belong to the set $(\tilde{a}\to\tilde{c})\wedge(\tilde{b}\to\tilde{d})$, which means that $p=\langle q,r\rangle$ where q maps $\tilde{a}$ to $\tilde{c}$ and r maps $\tilde{b}$ to $\tilde{d}$. Consider the set S, consisting of all pairs of different intersecting lines $\langle a,b\rangle$ such that q maps a to the common point of a and b, and r maps b to the common plane of a and b. Given q, r, and n we can generate all elements of S. As the pair $\langle\tilde{a},\tilde{b}\rangle$ belongs to S, we conclude that

$$7n-O(1)\leqslant K(\tilde{a},\tilde{b}|n)\leqslant K(p)+\log|S|+O(\log\log|S|).$$

Thus it suffices to prove that the cardinality of S does not exceed $2^{5.5n+O(1)}$. This is a direct corollary of the following lemma.

Lemma 1. *Let f be a function mapping every line to a point on that line, and g a function that maps every line to a plane containing that line. Let S consist of all pairs $\langle a,b\rangle$ such that $f(a)\in b$ and $a\subset g(b)$. Then $|S|\leqslant 2^{5.5n+O(1)}$.*

Proof. Let us see first what bound can be proven by easy arguments. For any line b there are at most $2^{2n+o(1)}$ lines a in the plane $g(b)$, hence the cardinality of S is at most $2^{2n+o(1)}$ times bigger than the number of lines ($2^{4n+o(1)}$); this gives the bound $|S|\leqslant 2^{6n+o(1)}$. The same bound can be proven by counting for every line a the number of lines b passing through $f(a)$. Note that in the first argument we did not use the fact that the line b passes through $f(a)$, and in the second one that the line a lies on the plane $g(b)$. Our plan is as follows: we will modify the first argument to show that the average number of pairs $\langle a,b\rangle$ in S having the same second component b is at most $2^{1.5n+O(1)}$. In that argument we will take into account the condition $f(a)\in b$. (We could argue in a symmetrical way to show that the average number of pairs $\langle a,b\rangle$ in S having the same first component a is at most $2^{1.5n+O(1)}$.)

Partition S into slices, every slice consists of all pair $\langle a,b\rangle\in S$ with the same value of $g(b)$. We will upperbound the number of pairs in every slice and then we will sum up the resulting bounds. So, fix a plane d and upperbound the number of $\langle a,b\rangle\in S$ such that $g(b)=d$.

To this end consider any point $c\in d$ and denote by A_c the set of all lines a on the plane d for which $f(a)=c$ and by B_c the set of all lines b passing through c for which $g(b)=d$. The number of pairs $\langle a,b\rangle\in S$ with $g(b)=d$ does not exceed

$$\sum_{c\in d}|A_c||B_c|\leqslant\sqrt{\sum_{c\in d}|A_c|^2\sum_{c\in d}|B_c|^2}.$$

Both sums in the right hand side have a clear interpretation. Indeed, $\sum_c|A_c|^2$ determines the probability of the following event: two lines a',a'' in the plane

d chosen at random satisfy $f(a') = f(a'')$. More specifically, let N denote the total number of lines on the plane d. Then

$$\begin{aligned}\operatorname{Prob}[f(a') = f(a'')] &= \sum_c \operatorname{Prob}[f(a') = f(a'') = c] \\ &= \sum_c \operatorname{Prob}[f(a') = c] \operatorname{Prob}[f(a'') = c] = \sum_c |A_c|^2/N^2.\end{aligned}$$

For any fixed a' the probability of event $f(a') = f(a'')$ does not exceed the probability that the line a'' passes through the point $f(a')$. The latter probability is equal to $2^{-n+o(1)}$, hence

$$\operatorname{Prob}[f(a') = f(a'')] \leqslant 2^{-n+o(1)} \implies \sum_c |A_c|^2 \leqslant N^2 2^{-n+o(1)} = 2^{3n+o(1)}.$$

The other sum $\sum_c |B_c|^2$ determines the average number of common points of two lines b', b'' chosen at random in the set M_d, consisting of those lines b for which $g(b) = d$ (all they lie in the plane d). More specifically,

$$\begin{aligned}\mathbf{E}\,|b' \cap b''| &= \sum_c \operatorname{Prob}[c \in b' \cap b''] \\ &= \sum_c \operatorname{Prob}[c \in b'] \operatorname{Prob}[c \in b''] = \sum_c |B_c|^2/|M_d|^2.\end{aligned}$$

Any two distinct lines have at most 1 common point, and equal lines have 2^n common points, therefore

$$\mathbf{E}\,|b' \cap b''| \leqslant 1 + 2^n \operatorname{Prob}[b' = b''] = 1 + 2^n/|M_d|,$$

hence

$$\sum_c |B_c|^2 = \mathbf{E}\,|b' \cap b''||M_d|^2 \leqslant |M_d|^2 + |M_d|2^n.$$

Recall that the number of those $\langle a, b\rangle \in S$ with $g(b) = d$ does not exceed $\sqrt{\sum_c |A_c|^2 \sum_c |B_c|^2}$, therefore it does not exceed

$$\sqrt{2^{3n+o(1)}(|M_d|^2 + |M_d|2^n)} \leqslant 2^{1.5n+o(1)}(|M_d| + 2^n)$$

(the latter inequality is proven by mere squaring). It remains to sum up the resulting bounds over d:

$$|S| \leqslant 2^{1.5n+o(1)} \sum_d (|M_d| + 2^n).$$

The families of lines M_d form a partition of the set of all lines, therefore the sum of their cardinalities is equal to $2^{4n+o(1)}$. The number of d's is equal to $2^{3n+o(1)}$, thus the sum over all d of 2^n is also equal to $2^{4n+o(1)}$. Hence

$$|S| \leqslant 2^{1.5n+o(1)}(2^{4n+o(1)} + 2^{4n+o(1)}) = 2^{5.5n+1+o(1)}. \quad \square$$

Thus the first inequality of Theorem 4 is proved. The second inequality also follows from Lemma 1. Let ρ be any random variable with $H(\tilde{\gamma}|\rho,\tilde{\alpha}) = H(\tilde{\delta}|\rho,\tilde{\beta}) = 0$. That is, for some functions F, G with probability 1 it holds that $\tilde{\gamma} = F(\rho,\tilde{\alpha})$, $\tilde{\delta} = G(\rho,\tilde{\beta})$. We have to show that $H(\rho) \geqslant 1.5n - O(1)$.

To this end fix a value p in the support of ρ and consider the functions $f(a) = F(p,a)$ and $g(b) = F(p,b)$. Given that $\rho = p$, with probability 1 it holds $\tilde{\gamma} = f(\tilde{\alpha})$, $\tilde{\delta} = g(\tilde{\beta})$. This means that both equalities hold for all pairs a, b of intersecting lines whose probability is positive (given that $\rho = p$). By Lemma 1 there are at most $2^{5.5n+O(1)}$ such pairs. As Shannon entropy does not exceed the logarithm of the number of possible outcomes, we can conclude that $H(\tilde{\alpha},\tilde{\beta}|\rho = p) \leqslant 5.5n + O(1)$. Therefore $H(\tilde{\alpha},\tilde{\beta}|\rho)$ does exceed $5.5n + O(1)$ as well. Since $\langle\tilde{\alpha},\tilde{\beta}\rangle$ is uniformly distributed, this implies that

$$7n - O(1) \leqslant H(\tilde{\alpha},\tilde{\beta}) \leqslant H(\tilde{\alpha},\tilde{\beta}|\rho) + H(\rho) \leqslant 5.5n + O(1) + H(\rho).$$

The theorem is proved. □

Note that the depth of expression $(p \to q) \wedge (r \to s)$ is equal to 2. The depth of all other examples above is also at most 2. There is another expression of depth 2 whose complexity is not expressible in terms of the complexity vector of involved strings, namely $(p \vee q) \to (r \vee s)$. A solution to this problem is a program that given any of two strings $0p, 1q$ computes any of strings r, s. Make in this formula the following substitution $p = a$, $q = b$, $r = \langle a, c\rangle$, $s = \langle b, d\rangle$ where $\langle a, b, c, d\rangle$ is one of the two quadruples constructed above. As given a it is much easier to compute $\langle a, c\rangle$ than $\langle b, d\rangle$, and given b it is much easier to compute $\langle b, d\rangle$ than $\langle a, c\rangle$, it is easy to show that the complexity of the resulting set is the same as that of $(a \to \langle a, c\rangle) \wedge (b \to \langle b, d\rangle)$. The latter problem is equivalent to the problem $(a \to c) \wedge (b \to d)$. And we know that the complexity of this problem is different for two above constructed quadruples.

One may prove by exhaustive search that the complexity of all other expressions of depth 2 is determined by the complexity vector of involved strings.

2.2 Strengthening Theorem 4

It is easy to see that for the quadruple $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$ used in the proof of Theorem 4, the upper bound of $K((\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d}))$ from Theorem 3 is equal to $2n$. And the established lower bound of $K((\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d}))$ is only $1.5n$. Thus we have a gap of $0.5n$ between proven lower and upper bounds. Are there a, b, c, d for which $K((a \to c) \wedge (b \to d))$ is equal to the upper bound of Theorem 3 while that upper bound is bigger than the lower bound of that theorem? The following theorem answers this question in positive.

Theorem 5. *For all n there are strings $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$ of complexity n such that the complexity of all pairs $\langle\tilde{a},\tilde{b}\rangle$, $\langle\tilde{a},\tilde{c}\rangle$, ..., $\langle\tilde{c},\tilde{d}\rangle$ is equal to $2n$, the complexity of all triples $\langle\tilde{a},\tilde{b},\tilde{c}\rangle$, $\langle\tilde{a},\tilde{b},\tilde{d}\rangle$, ... is equal to $3n$, the complexity of the quadruple $\langle\tilde{a},\tilde{b},\tilde{c},\tilde{d}\rangle$ is also equal to $3n$, and the complexity of the problem $(\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})$ is $2n$ (all the equalities hold up an additive $O(\log n)$ term).*

The similar statement holds for Shannon entropy.

It is easy to verify that for such quadruple the lower and upper bounds of Theorem 3 are equal to n and $2n$, respectively. Thus we indeed get an example we have been looking for.

Proof. Fix n. Call a set S of strings of length n *uniform* if for every triple $\langle a, b, c\rangle$ of strings of length n there is unique d such that $\langle a, b, c, d\rangle \in S$. For every n we will define a uniform set S that can be effectively found given n. As $\langle \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}\rangle$ we will take a random quadruple in S, that is, a quadruple in S such that $K(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}|n) \geqslant 3n$. (Such quadruple does exist, as the number of descriptions of length less than $3n$ equals $1 + 2 + \cdots + 2^{3n-1} < 2^{3n}$.) This will imply that $K(\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}) = K(\tilde{a}, \tilde{b}, \tilde{c}) = 3n$, and, as a corollary, $K(\tilde{a}) = K(\tilde{b}) = K(\tilde{c}) = n$, $K(\tilde{a}, \tilde{b}) = K(\tilde{a}, \tilde{c}) = K(\tilde{b}, \tilde{c}) = 2n$. To obtain the inequality $K((\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})) \geqslant 2n$, we will define S so that this inequality be true for most quadruples $\langle a, b, c, d\rangle$ in S (more specifically, the fraction of quadruples that do not satisfy the inequality will be $O(1/n)$). Then this inequality will be true for any random quadruple in S (for large enough n). To satisfy the remaining requirements on the complexity vector of $\langle \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}\rangle$ it suffices to ensure that both triples $\langle \tilde{a}, \tilde{b}, \tilde{d}\rangle$ and $\langle \tilde{a}, \tilde{c}, \tilde{d}\rangle$ have also complexity $3n$. Note that we need not care that $K(\langle \tilde{b}, \tilde{c}, \tilde{d}\rangle) = 3n$, since this is implied by the inequality $K((\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})) \geqslant 2n$. Indeed, we have

$$2n \leqslant K((\tilde{a} \to \tilde{c}) \wedge (\tilde{b} \to \tilde{d})) \leqslant K(\tilde{d}|\tilde{b}, \tilde{c}) + K(\tilde{c}) = K(\tilde{d}|\tilde{b}, \tilde{c}) + n \leqslant 2n,$$

hence

$$K(\tilde{d}|\tilde{b}, \tilde{c}) = n \Longrightarrow K(\tilde{b}, \tilde{c}, \tilde{d}) = K(\tilde{b}, \tilde{c}) + K(\tilde{d}|\tilde{b}, \tilde{c}) = 2n + n = 3n.$$

In the same way we will ensure that $K(\langle \tilde{a}, \tilde{b}, \tilde{d}\rangle) = K(\langle \tilde{a}, \tilde{c}, \tilde{d}\rangle) = 3n$. Namely, we will construct S so that the complexity of both problems $(c \to b) \wedge (a \to d)$ and $(b \to a) \wedge (c \to d)$ be also at least $2n$ for most quadruples in S. This implies via a symmetrical argument that both triples $\langle \tilde{a}, \tilde{b}, \tilde{d}\rangle$ and $\langle \tilde{a}, \tilde{c}, \tilde{d}\rangle$ are random for any random quadruple $\langle \tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}\rangle$ in S (for large enough n).

So let $k = O(\log n)$ where the constant in $O(\log n)$ is to be specified later. It suffices given n to find a uniform set S such that each of the three inequalities

$$\begin{aligned} K((a \to c) \wedge (b \to d)) &\geqslant 2n - k, \\ K((c \to b) \wedge (a \to d)) &\geqslant 2n - k, \\ K((b \to a) \wedge (c \to d)) &\geqslant 2n - k, \end{aligned} \tag{2}$$

holds for more than $(1 - O(1/n))$ fraction of quadruples in S.

The first inequality in (2) means that there is no $\langle p, q\rangle$ of complexity less than $2n - k$ such that $[p](a) = c$, $[q](b) = d$. So we have a small set M of pairs of functions (less than 2^{2n-k} pairs) and want to find a set S such that for most $\langle a, b, c, d\rangle$ in S there is no $\langle f, g\rangle \in M$ with $f(a) = c$ and $g(b) = d$. Unfortunately, we cannot compute M from n. Therefore we will construct S in such a way that for all small sets M of pairs of functions this property holds. In the following

lemma M denotes a set consisting of 2^{2n-k} pairs of total functions mapping binary strings of length n to binary strings of length n.

Lemma 2. *If $k = O(\log n)$ is appropriately chosen then for all n there is a uniform set S such that the following holds. For every M neither of the sets*

$$\{\langle a,b,c,d\rangle \in S \mid (\exists \langle f,g\rangle \in M)\ f(a)=c,\ g(b)=d\},$$
$$\{\langle a,b,c,d\rangle \in S \mid (\exists \langle f,g\rangle \in M)\ f(c)=b,\ g(a)=d\},$$
$$\{\langle a,b,c,d\rangle \in S \mid (\exists \langle f,g\rangle \in M)\ f(b)=a,\ g(c)=d\}$$

has more than $O(1/n)$ fraction of S.

Proof. We will prove that with high probability a uniform set S chosen at random satisfies the statement of the lemma. As the three above sets are symmetrical, it suffices to show that with probability close to 1 for every M the first of them has at most $O(2^{3n}/n)$ elements of S.

Say that M *serves* $\langle a,b,c,d\rangle$ if $f(a)=c$ and $g(b)=d$ for some pair $\langle f,g\rangle \in M$. Fix M and upperbound the probability that M serves more than $O(2^{3n}/n)$ of quadruples in S.

Call a triple $\langle a,b,c\rangle$ *bad* if $f(a)=c$ for more than $n2^{-n}|M|$ pairs $\langle f,g\rangle \in M$ (this property does not depend on b). Otherwise call the triple *good*. The probability of event $f(a)=c$ when $\langle a,b,c\rangle$ and $\langle f,g\rangle \in M$ are chosen at random is equal to 2^{-n}. Hence there are less than $2^{3n}/n$ bad triples.

For any good triple $\langle a,b,c\rangle$ the probability that M serves $\langle a,b,c,d\rangle$ for d chosen at random is at most $O(1/n)$. Indeed, if M serves $\langle a,b,c,d\rangle$ then d belongs to the set $\{g(b) \mid \langle f,g\rangle \in M,\ f(a)=c\}$. Since $\langle a,b,c\rangle$ is good, this set has less than $n2^{-n}|M| = 2^{\log n - n + 2n - k} < 2^n/n$ strings (provided $k > 2\log n$). Hence the probability that d falls into this set is $O(1/n)$.

We will use now the well known Chernoff bound [2]: assume that we make N independent trials and the probability of success in each trial is p; then for every $0 < \varepsilon \leqslant p(1-p)$ the fraction of successful trials is less than $p+\varepsilon$ with probability at least $1 - 2^{-\varepsilon^2 N/(2p)}$. For $\varepsilon = p/2$ Chernoff inequality implies that the probability that the number of successful trials is greater than $3p/2$ is less than $2^{-pN/8}$. In our case, trials correspond to good triples $\langle a,b,c\rangle$ and thus $N \geqslant 2^{3n-O(1)}$. In trial $\langle a,b,c\rangle$ we choose d at random and the trial is successful if the quadruple $\langle a,b,c,d\rangle$ is served by M. The success probability p is $O(1/n)$. Thus for a randomly chosen set S with probability at least $1 - 2^{-\Omega(N/n)}$ for at most $O(N/n)$ good triples $\langle a,b,c\rangle$ the quadruple $\langle a,b,c,d\rangle$ from S is served by M. The total number of quadruples in S served by M does not exceed the number of good served quadruples plus the total number of bad quadruples (a quadruples $\langle a,b,c,d\rangle$ is good if so is the triple $\langle a,b,c\rangle$). The first term is at most $O(N/n) = O(2^{3n}/n)$ with probability at least $1 - 2^{-\Omega(2^{3n}/n)}$. The second term is less than $2^{3n}/n$. Therefore with that probability the number of served quadruples in S is not greater than $2^{3n}/n + O(2^{3n}/n) = O(2^{3n}/n)$.

The number of different M's is at most

$$((2^n \cdot 2^n)^{2^n})^{2^{2n-k}} = 2^{2^{3n-k+\log 2n}}.$$

Hence with probability at least

$$1 - 2^{2^{3n-k+\log 2n} - 2^{3n-\log n - O(1)}}$$

every M serves at most $O(2^{3n}/n)$ of quadruples in S. If $k > 3\log n + O(1)$ this probability tends to 1 as n tends to infinity. □

The requirements on S in the lemma are decidable. Therefore given n we can find by brute force search the first set S satisfying the lemma. As explained above, any random quadruple from S satisfies the statement of the theorem.

It remains to handle the Shannon entropy case. Pick any set S, satisfying Lemma 2. Let $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}$ be independent random variables distributed uniformly among binary stings of length n, and let $\tilde{\delta}$ be the unique string d such that the quadruple $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, d$ is in S.

Let us show that $H((\tilde{\alpha} \to \tilde{\gamma}) \wedge (\tilde{\beta} \to \tilde{\delta})) \geqslant 2n - O(\log n)$. Assume that a random variable ρ is given such that $H(\tilde{\gamma}|\rho, \tilde{\alpha}) = H(\tilde{\delta}|\rho, \tilde{\beta}) = 0$. This means that there are functions F, G such that $\tilde{\gamma} = F(\rho, \tilde{\alpha})$, $\tilde{\delta} = G(\rho, \tilde{\beta})$ with probability 1. We have to prove that in this case $H(\rho) \geqslant 2n - O(\log n)$.

Recall that $H(\rho)$ is the average negative logarithm of the probability of the outcome p over all outcomes p of ρ. Therefore it suffices to show that only a small fraction ρ's outcomes have probability more more than 2^{-2n+k} (we will call such outcomes *heavy* and remaining outcomes *light*). More specifically, we will show that the cumulative probability of heavy outcomes is $O(1/n)$. This will imply the inequality

$$H(\rho) \geqslant (1 - O(1/n))(2n - k) = (1 - O(1/n))(2n - O(\log n)) \geqslant 2n - O(\log n).$$

To this end fix an outcome p in the support of ρ and consider the functions $f_p(a) = F(p, a)$ and $g_p(b) = F(p, b)$. Given that $\rho = p$, with probability 1 we have $\tilde{\gamma} = f(\tilde{\alpha})$, $\tilde{\delta} = g(\tilde{\beta})$. This means that both inequalities hold for all a, b, c, d in S that have positive probability given that $\rho = p$.

The number of heavy outcomes is at most 2^{2n-k}, since their cumulative probability cannot exceed 1. Let M consist of all pairs (f_p, g_p) for heavy p. By the property of S for a fraction at most $O(1/n)$ of quadruples a, b, c, d in S there is a heavy p such that the conditional probability of the quintuple a, b, c, d, p is positive. Hence the cumulative probability of all heavy outcomes is $O(1/n)$, q.e.d.

In a similar way we can show that $H((\tilde{\gamma} \to \tilde{\beta}) \wedge (\tilde{\alpha} \to \tilde{\delta})) \geqslant 2n - O(\log n)$ and $H((\tilde{\beta} \to \tilde{\alpha}) \wedge (\tilde{\gamma} \to \tilde{\delta})) \geqslant 2n - O(\log n)$. These inequalities imply exactly in the same was as in Kolmogorov complexity case that the quadruple $\tilde{\alpha}, \tilde{\beta}, \tilde{\gamma}, \tilde{\delta}$ has the desired entropy vector. □

There is another quadruple that has the same complexity vector and such that $K((a \to c) \wedge (b \to d)) = n$. Namely, pick a random string of length $3n$ and cut it into three parts, each of length n. In this way we obtain a, b and c; let then $d = a \oplus b \oplus c$. Given $a \oplus c$, we can transform a to c and b to d, Hence $K((a \to c) \wedge (b \to d)) = n$. Thus we have another proof of Theorem 4.

It is instructive to compare the new proof to the old one. In both proofs we show that there is a quadruple having some specific properties, namely, a quadruple having a given complexity vector and a given lower bound for $(a \to b) \wedge (c \to d)$. The lower bound in the second proof is stronger, however the proof itself is less constructive. In both proofs given n we effectively find a set S, and then pick a random element in S. The important difference is that in the first proof the set S is explicitly presented: it is the set of quadruples (a pair of intersecting lines, their common point, their common plane). In the second proof the set S is not explicitly presented. The proofs of the first type are called *effective*, the proofs of the second type are called *quasi-effective* (we consider only proofs of the existence of an object with certain property). In both cases there is a probabilistic algorithm that given n with high probability generates an object having the desired property. But in the first case its running time is bounded by a polynomial in n (addition, multiplication and division in the field F_n can be performed in polynomial time), while in the second case we do not know any efficient algorithm. It would be interesting to find out whether there is a set S satisfying Lemma 2 such that there is a polynomial time algorithm that given a, b, c finds d for which $\langle a, b, c, d\rangle \in S$.

Effective and quasi-effective proofs are opposed to non-effective ones—those in which we do not construct any algorithm to generate objects with the desired property. Usually, such a proof is easier to find than a quasi-effective one. The present paper is an exception: we do not know any easier proof of the existence of a, b, c, d for which the complexity of the problem $(a \to c) \wedge (b \to d)$ is greater than the lower bound of Theorem 3.

In conclusion we present another example of a theorem whose quasi-effective proof is at least as easy as the known non-effective ones.

3 Constructing strings having large amount of mutual information but have nothing in common

The paper [3] shows that for all n there are strings a_n, b_n such that $K(a_n) = K(b_n) = 2n$, $K(a_n, b_n) = 3n$ and hence a_n and b_n have $K(a_n) + K(b_n) - K(a_n, b_n) = n$ bits of mutual information. At the same time a_n, b_n have nothing in common in the following sense. Les us look for a word c of small complexity such that the conditional complexities of a and b are low when c is known. More specifically, consider the set of all those triples $\langle u, v, w\rangle$ of non-negative reals for which for all n there is a string c_n such that $K(c_n) \leqslant un + O(\log n)$, $K(a_n|c_n) \leqslant vn + O(\log n)$ and $K(b_n|c_n) \leqslant wn + O(\log n)$. This set is called the *positive complexity profile* of a and b and is denoted by M^+_{ab}. The larger this set is the more a, b have in common.

For instance, let (a, b) be a random pair of strings of length $2n$ that have common prefix of length n. Then M^+_{ab} contains the triple $\langle 1, 1, 1\rangle$ as we may let c be the common prefix of a and b. This pair has the largest M^+_{ab} among all

pairs (a, b) with $K(a) = K(b) = 2n$, $K(a, b) = 3n$.

We say that a and b have nothing in common, if their complexity profile is minimal among complexity profiles of pairs with $K(a) = K(b) = 2n$, $K(a, b) = 3n$. It turns out that the minimal complexity profile consists of all triples $\langle u, v, w\rangle$ satisfying the inequalities

$$u + v \geqslant 2, \quad u + w \geqslant 2, \quad u + v + w \geqslant 3$$

and **at least one** of the inequalities

$$u + v \geqslant 3, \quad u + w \geqslant 3, \quad u + v + w \geqslant 4.$$

Let $M^+_{\min}$ denote this set. It is easy to verify that the profile of any a, b such that $K(a) = K(b) = 2n$, $K(a, b) = 3n$ includes the set $M^+_{\min}$. To show this it suffices to consider only strings c composed of some substrings of the shortest programs that print a and b and of the shortest programs mapping a to b and b to a. (For detailed proof see [3].) A result of [3] states that the there are pairs of strings whose complexity profile equals $M^+_{\min}$:

Theorem 6 ([3])**.** *There are sequences a_n, b_n such that $K(a_n) = K(b_n) = 2n$, $K(a_n, b_n) = 3n$ whose profile is equal to M^+_{min} (all equalities are valid up to an additive $O(\log n)$ term).*

The proof of this theorem presented in the paper [3] is non-effective and its authors wonder whether there is a quasi-effective one.[1]

A quasi-effective proof of Theorem 6. The profile of a, b is equal to $M_{\min}$ if the following holds. Any triple u, v, w such that $u + v < 2$, or $u + w < 2$, or $u + v + w < 3$, or simultaneously

$$u + v < 3, \quad u + w < 3, \quad u + v + w < 4$$

does not belong to the profile of a, b. If at least one of the inequalities $u+v < 2$, $u + w < 2$, $u + v + w < 3$ holds then the profile of all a, b with $K(a_n) = K(b_n) = 2n$, $K(a_n, b_n) = 3n$ does not contain $\langle u, v, w\rangle$. This is a direct corollary of the inequalities $K(a) \leqslant K(c) + K(a|c)$, $K(b) \leqslant K(c) + K(b|c)$, $K(a, b) \leqslant K(c) + K(a|c) + K(b|c)$, respectively. So it suffices to construct strings a_n, b_n for which $K(a_n, b_n) = 3n + O(\log n)$, $K(a_n) = K(b_n) = 2n + O(\log n)$ and that have the following property. For any U, V, W satisfying the inequalities

$$\begin{aligned} U + \max\{V, W\} &< 3n - k, \\ U + V + W &< 4n - k, \end{aligned} \tag{3}$$

[1] Let us mention that the complexities of a, b and of the pair $\langle a, b\rangle$ in Theorem 6 are chosen to be $2n, 2n, 3n$ to simplify the statement. A similar result holds for strings with any fixed proportions between complexities of a, b and of their pair (of course the set $M^+_{\min}$ depends on the chosen proportions). By the same reason we restrict our attention to pairs with $K(a_n) = K(b_n) = 2n$, $K(a_n, b_n) = 3n$.

there is no c such that

$$K(c) < U, \quad K(a_n|c) < V, \quad K(b_n|c) < W.$$

Here k is a linear function of $\log n$ to be specified later. For every n we will construct a directed graph whose vertices are strings of length $2n$ and the number of edges is $2^{3n+O(1)}$. As $\langle a, b\rangle$ we will take any random edge in the graph. Then $K(a, b)$ will be about $3n$. We need also that the graph have two other properties. The first one is the following. For any set N of less than 2^{2n-k} vertices, only a small fraction of edges are incident to a node in N. Applying this property to the set of strings of complexity less than $2n - k$, we will show that for any random edge $\langle a, b\rangle$ of the graph both $K(a)$ and $K(b)$ are greater than $2n - k = 2n - O(\log n)$. The other property is the following. Most edges of the graph should not belong to the set

$$\{\langle a, b\rangle \mid \text{there are } U, V, W, \text{ satisfying inequalities (3) and a string } c \text{ such that } K(c) < U, \quad K(a_n|c) < V, \quad K(b_n|c) < W\}.$$

Since we are unable to compute this set given n, we have to find its decidable property such that every set with this property has few edges of the graph. Here is this property. Our set is a union over all U, V, W, satisfying inequalities (3), of some sets M_{UVW} where each M_{UVW} is a union of 2^U sets of the form $A \times B$ where $|A| < 2^V$, $|B| < 2^W$. So let us call any set of pairs of strings of length $2n$ *special* if it has this property. Thus it suffices to prove the following lemma.

Lemma 3. *For some function $k = k_n = O(\log n)$ and for all n there is a directed graph whose nodes are strings of length $2n$ that has the following properties. 1. The number of edges is between 2^{3n-2} and 2^{3n}. 2. For any set M of cardinality less than 2^{2n-k} at most $O(2^{3n}/n)$ edges are incident to a node in M. 3. Every special set of pairs of strings of length $2n$ has at most $O(2^{3n}/n)$ edges of the graph.*

Proof. We will show that a randomly chosen graph has these properties with high probability. The probability distribution over graphs is defined as follows. We make 2^{3n} independent trials choosing in each trial an edge at random (we allow loops). Let us prove that with high probability all the three properties hold.

1. For any fixed set of 2^{3n-2} pairs of strings the probability that a random edge falls into the set is 2^{-n-2}. The probability that all 2^{3n} edges fall into the set is $2^{-(n+2)2^{3n}}$. The number of such sets is less than $2^{4n\cdot 2^{3n-2}} = 2^{n\cdot 2^{3n}}$. Thus the probability that all edges get into a set of cardinality 2^{3n-2} does not exceed $2^{-(n+2)2^{3n}+n\cdot 2^{3n}} = 2^{-2\cdot 2^{3n}} \ll 1$.

2. The probability that a random edge in incident to a node in a fixed set M of cardinality 2^{2n-k} is less than $1/n$ (provided $k > \log n$). By Chernoff bound the probability that this happens for a fraction at least $2/n$ of 2^{3n} random edges does not exceed $2^{-\Omega(2^{3n}/n)} = 2^{-2^{3n-\log n-O(1)}}$. The number of such M's is less

than $2^{2^{2n}}$. Hence with probability at least

$$1 - 2^{2^{2n} - 2^{3n - \log n - O(1)}} \approx 1$$

for every M at most $O(2^{3n}/n)$ edges of the graph are incident to a node in M.

3. Let us prove that any special set has at most $O(2^{4n}/n)$ edges. The number of U, V, W does not exceed $O(n^3)$. For fixed U, V, W any union of 2^U sets of the form $A \times B$ where $|A| < 2^V$, $|B| < 2^W$ has at most $2^{U+V+W} \leqslant 2^{4n-k}$ pairs. Multiplying the latter number by $O(n^3)$ we get $2^{4n-k+3\log n+O(1)} = O(2^{4n}/n)$ (provided $k > 4\log n$). Therefore for any special M with probability at least $1 - O(1/n)$ a random edge does not fall into M. By Chernoff bound with probability $1 - 2^{-\Omega(2^{3n}/n)}$ at most $O(2^{3n}/n)$ edges of the graph fall into M.

Let us estimate the number of special sets. For any fixed U, V, W there are at most

$$(2^{2n \cdot 2^V} \cdot 2^{2n \cdot 2^W})^{2^U} = 2^{2n(2^V + 2^W)2^U} < 2^{2^{3n-k+\log n+O(1)}}$$

special sets. Rising this number to the power of the number of different U, V, W we obtain an upper bound $2^{2^{3n-k+4\log n+O(1)}}$ for the number of special sets. Recall that any special set has more than $\Omega(2^{3n}/n)$ edges with probability at most $2^{-2^{3n-\log n-O(1)}}$. If $k > 5\log n + O(1)$ then even multiplied by the number of special sets this probability is less than 1. Thus with high probability all special sets have less than $O(2^{3n}/n)$ edges. □

Acknowledgement. The authors are grateful to A. Shen for the question.

References

[1] C.H. Bennett, P. Gács, M. Li, P. Vitányi and W. Zurek. Information Distance. *IEEE transactions on Information Theory*, Vol. 44 (1998) No. 4, 1407–1423.

[2] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23 (1952) 493–509.

[3] An. Muchnik, A. Romashchenko, A. Shen, and N. Vereshchagin. Upper semilattice of binary strings with the relation "x is simple conditional to y". *Theoretical Computer Science* 271 (2002) 69–95.

[4] Muchnik An. A. Conditional complexity and codes, *Theoretical Computer Science*, v. 271 (2002) p. 97–109.

[5] An. Muchnik and N. Vereshchagin. "Shannon Entropy vs. Kolmogorov Complexity". Computer Science — Theory and Applications: First International Computer Science Symposium in Russia, CSR 2006, St. Petersburg, Russia, June 8-12. 2006. Proceedings. Editors: Dima Grigoriev, John Harrison, Edward A. Hirsch. Lecture Notes in Computer Science, vol. 3967 (2006) 281–291.

[6] A. Shen, N. Vereshchagin. Logical operations and Kolmogorov complexity. *Theoretical Computer Science* 271 (2002) 125–129..

[7] A.K. Zvonkin, L.A. Levin, The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russian Math. Surveys*, v. 25 (1970) no. 6, 83–124.