ELSEVIER

Theoretical Computer Science 271 (2002) 69–95

Theoretical Computer Science

www.elsevier.com/locate/tcs

Upper semi-lattice of binary strings with the relation "x is simple conditional to y"

Alexei Chernov[a,*], Andrej Muchnik[b,1], Andrei Romashchenko[a],
Alexander Shen[c,2], Nikolai Vereshchagin[a,3]

[a] *Department of Mathematical Logic and Theory of Algorithms, Moscow State University, Vorobjewy Gory, Moscow, Russia 119899*
[b] *Institute of New Technologies of Education, 10 Nizhnyaya Radischewskaya, Moscow, Russia 109004*
[c] *Institute of Problems of Information Transmission Russia*

Abstract

In this paper we construct a structure R that is a "finite version" of the semi-lattice of Turing degrees. Its elements are strings (technically, sequences of strings) and $x \leqslant y$ means that $K(x|y)=$ (conditional Kolmogorov complexity of x relative to y) is small. We construct two elements in R that do not have greatest lower bound. We give a series of examples that show how natural algebraic constructions give two elements that have lower bound 0 (minimal element) but significant mutual information. (A first example of that kind was constructed by Gács–Körner (Problems Control Inform. Theory 2 (1973) 149) using a completely different technique.) We define a notion of "complexity profile" of the pair of elements of R and give (exact) upper and lower bounds for it in a particular case. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Kolmogorov complexity; Common information; Conditional complexity

1. Introduction

Let α and β be two infinite binary sequences. We say that α is Turing reducible to β if there exists a Turing machine M that produces α on its output tape when β is provided on input tape. Turing reducibility is reflexive and transitive, so we get a pre-order on the set of all infinite binary sequences (this pre-order is usually denoted by $\leqslant_T$). The equivalence classes $((x \sim y) \Leftrightarrow (x \leqslant_T y) \wedge (y \leqslant_T x))$ form an upper

[*] Corresponding author.

E-mail addresses: an@romash.mccme.ru (Andrei Romashchenko), shen@mccme.ru (Alexander Shen), ver@mech.math.msu.su (Nikolai Vereshchagin).

[1] Fax: +709-59-15-6963.

[2] The work was supported by Volkswagen Foundation while visiting Bonn University.

[3] The work was partially done while visiting the University of Amsterdam and DIMACS center.

0304-3975/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0304-3975(01)00032-9

semi-lattice whose elements are called Turing degrees. This semi-lattice is well studied in recursion theory (see, e.g., [7])

Now let us replace infinite sequences α and β by finite binary strings x and y. Of course, for any x and y there exists a Turing machine M that produces x from y. So to get a non-trivial relation we have to put some restrictions on M. It is natural to require that M is simple (its program is short compared to x and y). Here the notion of Kolmogorov complexity comes into play. By definition, the conditional Kolmogorov complexity $K(x|y)$ is the length of the shortest program that produces x having y as an input. Now we can define the relation $x \leqslant_c y$ as $K(x|y) \leqslant c$ (here x and y are binary strings, c is a number).

If c is a constant, this relation does not have good properties (for example, it is not transitive). This relation also depends on a specific programming language used in the definition of Kolmogorov complexity. To overcome these difficulties, we use the standard trick and consider the asymptotic behavior of the complexity for sequences of strings.

Let $\boldsymbol{x} = x_1, x_2, \ldots$ be a sequence of binary strings. We call it *regular* if length of x_i is polynomially bounded, i.e., if $|x_i| \leqslant ci^k$ for some c, k and for all i. Let R denote the set of all regular sequences. We say that regular sequence $\boldsymbol{x}$ is *simple* conditional to a regular sequence $\boldsymbol{y}$ if

$$K(x_i|y_i) = \mathrm{O}(\log i)$$

and write $\boldsymbol{x} \leqslant \boldsymbol{y}$. The $\leqslant$-relation is a pre-order defined on R. The relation $(x \leqslant y) \wedge (y \leqslant x)$ is an equivalence relation. Equivalence classes form a partially ordered set which (for the same reasons as in the case of Turing degrees) is an upper semi-lattice (any two elements have a least upper bound).

We prove (Section 2) that this set is not a lower semi-lattice: there are two elements that do not have greatest lower bound. Note that the set of Turing degrees is not a lower semi-lattice either (see, e.g., [7]), but our proof goes in a completely different way.

The semi-lattice R is useful for analyzing the notion of common information. This notion was introduced by Gács and Körner [1] in the context of Shannon information theory. They also described a similar notion in the algorithmic theory but do not give a precise definition. We give such a definition in terms of the semi-lattice R (Section 3).

The main result of [1] is an example of two objects whose “common information” is far less than their “mutual information”; Gács and Körner provide such an example in context of Shannon information theory and mention that it could be reformulated for algorithmic information theory. This example was analyzed in [2] where an alternative proof for a special case of Gács–Körner example was provided.

A completely different example of two strings whose common information is much less than their mutual information was given in [4]; for details see [5].

In this paper we develop a third approach to construct such pairs of strings. It is based on the geometry of finite fields. Several examples of this type are given in

Section 4. Our examples (as well as Gács–Körner's) are constructive in the following sense. In the recursion theory, we call a proof of a theorem of the form $\forall n \exists a\ P(n,a)$ constructive if there exists an algorithm that given n computes an object a_n such that $P(n,a_n)$. In our context this makes no sense, as in this case the complexity of a_n is bounded by $\log n$ and we are interested in properties $P(n,a)$ implying that complexity of a is linear in n. We find reasonable the following meaning of constructivity here: there is a probabilistic algorithm that given n with high probability outputs such an object a that $P(n,a)$. More specifically, the probability should tend to 1 as n tends to infinity. All our examples except one from Theorem 7(c) are constructive in this sense.

The amount of common information does not determine completely how much the strings x and y have in common. What reflects this better is the "complexity profile of x and y", defined as the set of triples (u,v,w) such that $K(z) \leqslant u$, $K(x|z) \leqslant v$, and $K(y|z) \leqslant w$ for some string z. We use the method of [5] to find exact upper and lower bounds for complexity profile (Section 6). (Technically we have to speak not about strings x and y but about sequences of strings $x_0, x_1, \ldots$ and $y_0, y_1, \ldots$ such that complexity of x_i and y_i is proportional to i; see Section 6 for details.)

2. The upper semi-lattice $\boldsymbol{R}$

Let us recall the definition of conditional Kolmogorov complexity. Let U be a computable (partial) function of two arguments; arguments and values are binary strings. (Informally, U is an interpreter of some programming language, the first argument is a program and the second one is program's input.) Let us define $K_U(x|y)$ as $\min\{|p| : U(p,y) = x\}$; here $|p|$ stands for the length of p. There exists an optimal U, that is, a U such that $K_U \leqslant K_V + \mathrm{O}(1)$ for any other computable function V. We fix some optimal U and call $K_U(x|y)$ the *conditional complexity* of x when y is known.

The *unconditional* Kolmogorov complexity can be defined as $K(x|\Lambda)$ where Λ is the empty string. It turns out (see, e.g., [3]) that conditional complexity can be expressed in terms of unconditional complexity. Indeed, let us fix some computable bijection $p,q \mapsto \langle p,q \rangle$ between pairs of strings and strings. Then

$$K(\langle p,q \rangle) = K(p) + K(q|p) + \mathrm{O}(\log(|p| + |q|)).$$

A sequence $\boldsymbol{x} = x_1, x_2, \ldots$ of binary strings is called *regular* if there exist constants c and k such that $|x_i| \leqslant ci^k$ for all i. The set of all regular sequences is denoted by R. We define a pre-order on R saying that $\boldsymbol{x} = x_1, x_2, \ldots$ precedes $\boldsymbol{y} = y_1, y_2, \ldots$ if there exists a constant c such that $K(x_i|y_i) \leqslant c \log i$ for all i. (Let us agree that $\log x$ means $\log_2(x+2)$ so $\log x$ is positive for all $x \geqslant 0$ and we do not need to consider the case $i = 1$ separately.)

The O-term guarantees that the definition does not change if we replace the optimal function U used in the definition of Kolmogorov complexity by another optimal function. Moreover, since we use $\mathrm{O}(\log i)$ (and not $\mathrm{O}(1)$), the definition remains the same if we replace conditional Kolmogorov complexity defined as above by prefix complexity (see [3] for the definition). Indeed, these complexities differ only by $O(\log n)$ for strings of length n. Since elements of R are regular, this difference is absorbed by $O(\log i)$-term.

Two elements $\boldsymbol{x}$ and $\boldsymbol{y}$ are *equivalent* if $\boldsymbol{x} \leqslant \boldsymbol{y}$ and $\boldsymbol{y} \leqslant \boldsymbol{x}$. The equivalence classes form a partially ordered set. We denote this set by $\boldsymbol{R}$.

Proposition 1. *The set $\boldsymbol{R}$ is an upper semi-lattice: any two elements have a least upper bound.*

Proof. By definition, $\boldsymbol{z} \in \boldsymbol{R}$ is a least upper bound of $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{R}$ if

- $\boldsymbol{z}$ is an upper bound for $\boldsymbol{x}$ and $\boldsymbol{y}$, i.e., $\boldsymbol{x} \leqslant \boldsymbol{z}$ and $\boldsymbol{y} \leqslant \boldsymbol{z}$;
- $\boldsymbol{z} \leqslant \boldsymbol{u}$ for any other upper bound $\boldsymbol{u}$ of $\boldsymbol{x}$ and $\boldsymbol{y}$.

Let $\boldsymbol{x} = x_1, x_2, \ldots$ and $\boldsymbol{y} = y_1, y_2, \ldots$ be any two elements of $\boldsymbol{R}$. Consider the sequence $\boldsymbol{z} = z_1, z_2, \ldots$ where $z_i = \langle x_i, y_i \rangle$. (Recall that $p, q \mapsto \langle p, q \rangle$ denotes a computable bijection between pairs of strings and strings.) It is easy to see that $\boldsymbol{z}$ is regular and is the least upper bound for $\boldsymbol{x}$ and $\boldsymbol{y}$. □

Theorem 2. *The ordered set $\boldsymbol{R}$ is not a lower semi-lattice: there exist two elements $\boldsymbol{x}$ and $\boldsymbol{y}$ that do not have a greatest lower bound.*

Proof. To prove the theorem we have to construct two sequences $\boldsymbol{x}$ and $\boldsymbol{y}$ that have no greatest lower bound. Assume some n is fixed; let us explain how nth terms of $\boldsymbol{x}$ and $\boldsymbol{y}$ are constructed. Consider $2n$ binary strings of length n denoted by

$$b_1^0, b_2^0, \ldots, b_n^0, b_1^1, b_2^1, \ldots, b_n^1,$$

and one more string of length n denoted by

$$\varepsilon = \varepsilon_1 \ldots \varepsilon_n$$

(ε_i are individual bits). We want all these strings to be random and independent in the following sense: its concatenation is a string of length $2n^2 + n$ which is incompressible (its Kolmogorov complexity is equal to its length up to $O(1)$ additive term). Such strings do exist, see [3]. Now consider two strings

$$x = b_1^0 b_2^0 \ldots b_n^0 b_1^1 b_2^1 \ldots b_n^1$$

and

$$y = b_1^{\varepsilon_1} b_2^{\varepsilon_2} \ldots b_n^{\varepsilon_n}.$$

Strings x and y are nth terms of the sequences $\boldsymbol{x}$ and $\boldsymbol{y}$.

Let us mention that the pair $\langle x, y \rangle$ contains the same information as the concatenation string of length $2n^2 + n$ mentioned above, so the complexity of the pair $\langle x, y \rangle$ is $2n^2 + n + O(1)$. (As x is random, $b_i^0 \neq b_i^1$ for all i.)

In the sequel we use the following terminology. Strings b_i^e (for $e = 0, 1$ and $i = 1, \ldots, n$) are called *blocks*. We have $2n$ blocks; each block has length n. All the blocks

$b_i^{\varepsilon_i}$ that are included in y are called *selected* blocks; all other blocks $b_i^{1-\varepsilon_i}$ are called *omitted* blocks. Our construction starts with n pairs of blocks and a string ε that says which block is selected in each pair. The string x is a concatenation of all $2n$ blocks; the string y is a concatenation of n selected blocks.

Now the proof goes as follows. Each selected block is simple relative to both x and y since it is a substring of both x and y and position information could be encoded by $\mathrm{O}(\log n)$ bits. (When we say that a string u is *simple* relative to a string v we mean that $K(u|v)=\mathrm{O}(\log n)$.)

Suppose that the greatest lower bound of x and y exists. Let us denote it by z. Then any selected block is simple relative to z. On the other hand, any omitted block could not be simple relative to z. Indeed, assume that some omitted block b is simple relative to z. Then b is simple relative to y since z is simple relative to y by assumption. Then to restore x from y it is enough to specify the string ε and $n-1$ omitted blocks different from b, i.e., n^2 bits, and the complexity of pair $\langle x, y\rangle$ is at most $2n^2+\mathrm{O}(\log n)$ (n^2 bits in y and n^2 bits to specify x when y in known). This contradiction shows that no omitted block is simple relative to z.

Now let us show that y is simple relative to x. Indeed, to find y when x is known we need only to distinguish between omitted and selected blocks in each pair of blocks. We may assume that z is known since it is simple relative to x. Then we may enumerate all the objects that have small complexity relative to z until we find n blocks (we have the list of all blocks since we know x). These n blocks will be (as shown above) exactly the selected blocks, and we are done. So y is simple relative to x. But this is impossible, because in this case the pair $\langle x, y\rangle$ will have complexity at most $2n^2+\mathrm{O}(\log n)$ (instead of $2n^2+n$).

In the argument above we were quite vague about O-notation, so let us repeat the same argument more formally. The construction described above is performed for each n; to indicate the dependence on n let us write $x(n)$ instead of x, $b_i^0(n)$ instead of b_i^0, etc. Assume that $\boldsymbol{z}=z(0),z(1),\ldots$ is a greatest lower bound of $\boldsymbol{x}$ and $\boldsymbol{y}$. The first step in the proof is the following lemma.

Lemma 1. *There exists some constant c such that*

$$K(b|z(n))\leqslant c\ \log n$$

for any n and for any block b that was selected at nth step of the construction. (*There were n selected blocks at nth step*; *each of them has length n.*)

Indeed, consider all the blocks b that were selected at nth step; let $b(n)$ be one of them for which the complexity $K(b|z(n))$ is maximal. The sequence $\boldsymbol{b}=b(1),b(2),\ldots$ belongs to R. It is easy to see that $\boldsymbol{b}\leqslant\boldsymbol{x}$ and that $\boldsymbol{b}\leqslant\boldsymbol{y}$, because $b(n)$ is a substring of both $x(n)$ and $y(n)$. Therefore, $\boldsymbol{b}\leqslant\boldsymbol{z}$, since z is the greatest lower bound of $\boldsymbol{x}$ and $\boldsymbol{y}$. By definition,

$$K(b(n)|z(n))\leqslant c\ \log n$$

for some constant c; the same inequality is valid for all other selected blocks b since $b(n)$ has maximal complexity (relative to $z(n)$) among them. Lemma 1 is proved.

Lemma 2. *There exists some constant c such that*

$$K(b|y(n)) \geqslant n - c \log n$$

for any n and for any block b that was omitted at nth step of the construction.

Proof. As we have said, the string $x(n)$ can be reconstructed from the string $y(n)$, the string $\varepsilon(n)$, some omitted block b, its number and the concatenation of all other omitted blocks. Here all the information except b has bit size $n^2+n+(n^2-n)+\mathrm{O}(\log n)=2n^2+\mathrm{O}(\log n)$, and this information includes $y(n)$. Therefore, the complexity of $\langle x(n), y(n)\rangle$ does not exceed $K(b|y(n))+2n^2+\mathrm{O}(\log n)$. On the other hand, the complexity of $\langle x(n), y(n)\rangle$ is $2n^2+n+\mathrm{O}(1)$. Comparing the two inequalities, we see that $K(b|y(n)) \geqslant n-\mathrm{O}(\log n)$. Lemma 2 is proved. □

Lemma 3. *There exists some constant c such that*

$$K(b|z(n)) \geqslant n - c \log n$$

for any n and for any block b that was omitted at nth step of the construction.

Indeed, recall that $K(z(n)|y(n))=\mathrm{O}(\log n)$ by our assumption; note also that $K(b|y(n)) \leqslant K(b|z(n))+K(z(n)|y(n))+\mathrm{O}(\log n)$. Hence, $n-\mathrm{O}(\log n) \leqslant K(b|y(n)) \leqslant K(b|z(n))+K(z(n)|y(n))+\mathrm{O}(\log n)=K(b|z(n))+\mathrm{O}(\log n)$. Lemma 3 is proved.

Lemma 4. $K(\varepsilon(n)|x(n))=\mathrm{O}(\log n)$.

Proof. Lemma 1 implies that for big n the value $K(b|z(n))$ is less than $n/2$ for any selected block b; Lemma 3 implies that for big n the value $K(b|z(n))$ is bigger than $n/2$ for any omitted block b. Therefore, knowing $x(n)$ and $z(n)$ we can reconstruct the list of selected blocks just enumerating the strings s such that $K(s|z(n))<n/2$ until n blocks from $x(n)$ appear. Since $K(z(n)|x(n))=\mathrm{O}(\log n)$ by assumption, we need only $\mathrm{O}(\log n)$ additional bits to reconstruct $\varepsilon(n)$ from $x(n)$. Lemma 4 is proved. □

We conclude that $K(\langle x(n), \varepsilon(n)\rangle)$ is $2n^2+\mathrm{O}(\log n)$ but it should be $2n^2+n+\mathrm{O}(1)$. The contradiction shows that $\boldsymbol{x}$ and $\boldsymbol{y}$ do not have the greatest lower bound. □

Let us mention some other properties of the semi-lattice $\boldsymbol{R}$.

1. The operations "infimum" and "supremum" do not satisfy the distributive law even when they are defined. Indeed, consider sequences $\boldsymbol{x}$ and $\boldsymbol{y}$ where x_n and y_n are random independent strings of length n. Let $z_n=x_n \oplus y_n$ (bitwise addition modulo 2). Then

$$\sup(\inf(\boldsymbol{x},\boldsymbol{y}),\boldsymbol{z}) \neq \inf(\sup(\boldsymbol{x},\boldsymbol{z}),\sup(\boldsymbol{y},\boldsymbol{z})),$$

since $\inf(\boldsymbol{x},\boldsymbol{y})=\Lambda$ (where Λ is the least element of the semi-lattice), so the left-hand side is equal to $\boldsymbol{z}$ while the right-hand side is equal to $\sup(\boldsymbol{x},\boldsymbol{y})$.

Moreover,

$$\inf(\sup(\boldsymbol{x},\boldsymbol{y}),\boldsymbol{z}) \neq \sup(\inf(\boldsymbol{x},\boldsymbol{z}),\inf(\boldsymbol{y},\boldsymbol{z})),$$

since left-hand side is equal to $\boldsymbol{z}$ and right-hand side is equal to Λ.

2. For any two elements $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\boldsymbol{R}$ there exists a sequence $\boldsymbol{z}$ such that $\sup(\boldsymbol{y},\boldsymbol{z})=\sup(\boldsymbol{y},\boldsymbol{x})$ and $\inf(\boldsymbol{y},\boldsymbol{z})=\Lambda$. Indeed, given x,y and $K(x|y)$ we can enumerate the set of all programs p such that $p(y)=x$ and length of p is equal to $K(x|y)$. Let z be the first program in this enumeration.

This $\boldsymbol{z}$ could be considered as a "difference" between $\boldsymbol{x}$ and $\boldsymbol{y}$. Difference is not defined uniquely; for instance, if x_n and y_n are random independent strings of length n, both x_n and $x_n \oplus y_n$ are differences of x_n and y_n.

The semi-lattice $\boldsymbol{R}$ is only one of the possible refinements of the intuitive notion "x is simple relative to y". Here is another possibility. Let us fix a function $\log n \leqslant f(n)=o(n)$; assume that $\boldsymbol{x}$ and $\boldsymbol{y}$ are sequences of strings such that $|x_n|=O(n)$, $|y_n|=O(n)$. Define $\boldsymbol{x}\leqslant_f\boldsymbol{y}$ as $K(x_n|y_n)=O(f(n))$. One can show that this definition gives a semi-lattice with similar properties (no greatest lower bound; however, the proof is more difficult and is omitted).

3. Common and mutual information

The semi-lattice $\boldsymbol{R}$ is a useful tool to analyze the amount of common information shared by two strings.

Let x and y be two strings. By *mutual* information in x and y we mean the value $I(x:y)=K(x)+K(y)-K(\langle x,y\rangle)$. (Sometimes $I(x:y)$ is defined as $K(y)-K(y|x)$, but these quantities differ only by $O(\log n)$ for strings of length at most n, see [3].)

Theorem 3. *Let $\boldsymbol{x}=x_1,x_2,\ldots$ and $\boldsymbol{y}=y_1,y_2,\ldots$ be elements of $\boldsymbol{R}$.*

(a) *If $\boldsymbol{z}=z_1,z_2,\ldots$ is a lower bound of $\boldsymbol{x}$ and $\boldsymbol{y}$ then*

$$K(z_n)\leqslant I(x_n:y_n)+O(\log n). \tag{1}$$

(b) *If $\boldsymbol{z}=z_1,z_2,\ldots$ is a lower bound of $\boldsymbol{x}$ and $\boldsymbol{y}$ and*

$$K(z_n)=I(x_n:y_n)+O(\log n) \tag{2}$$

then $\boldsymbol{z}$ is the greatest lower bound of $\boldsymbol{x}$ and $\boldsymbol{y}$ in $\boldsymbol{R}$.

Proof. (a) Since $\boldsymbol{z}\leqslant\boldsymbol{x}$,

$$K(\langle x_n,z_n\rangle)=K(x_n)+K(z_n|x_n)+O(\log n)=K(x_n)+O(\log n).$$

So

$$K(x_n)=K(\langle x_n,z_n\rangle)+O(\log n)=K(z_n)+K(x_n|z_n)+O(\log n). \tag{3}$$

Similarly

$$K(y_n) = K(\langle y_n, z_n\rangle) + O(\log n) = K(z_n) + K(y_n|z_n) + O(\log n). \tag{4}$$

On the other hand,

$$K(\langle x_n, y_n\rangle) \leqslant K(z_n) + K(x_n|z_n) + K(y_n|z_n) + O(\log n). \tag{5}$$

since we can reconstruct the pair $\langle x_n, y_n\rangle$ from z_n and programs that transform z_n into x_n and y_n. Combining the last three inequalities [(3) + (4) − (5)], we get the statement (a).

Let us prove the part (b) now. Assume that $\boldsymbol{z}$ is a lower bound for $\boldsymbol{x}$ and $\boldsymbol{y}$ and inequality (1) turns into equality (2). Let $\boldsymbol{z}'$ be any other lower bound for $\boldsymbol{x}$ and $\boldsymbol{y}$. Consider the sequence $\boldsymbol{z}''$ defined as $z''_n = \langle z_n, z'_n\rangle$. It is the least upper bound of $\boldsymbol{z}$ and $\boldsymbol{z}'$ (Proposition 1). Therefore $\boldsymbol{z}'' \leqslant \boldsymbol{x}$ and $\boldsymbol{z}'' \leqslant \boldsymbol{y}$. Applying (a) to $\boldsymbol{z}''$ we see that

$$K(z''_n) = K(\langle z_n, z'_n\rangle) \leqslant I(x_n : y_n) + O(\log n)$$

By assumption, $I(x_n : y_n) = K(z_n) + O(\log n)$, so $K(\langle z_n, z'_n\rangle) \leqslant K(z_n) + O(\log n)$. On the other hand, $K(\langle z_n, z'_n\rangle) = K(z_n) + K(z'_n|z_n) + O(\log n)$, therefore $K(z'_n|z_n) \leqslant O(\log n)$ and $\boldsymbol{z}' \leqslant \boldsymbol{z}$ in $\boldsymbol{R}$. □

Remark. If two sequences $\boldsymbol{x} = x_1, x_2, \ldots$ and $\boldsymbol{y} = y_1, y_2, \ldots$ have the greatest lower bound $\boldsymbol{z} = z_1, z_2, \ldots$, one may call $K(z_n)$ “the amount of common information in strings x_n and y_n”.

4. Examples where common information is less than mutual information

Informally speaking, strings x and y have u-bit common information z if $K(z) = u$, $K(z|x) \approx 0$, and $K(z|y) \approx 0$. We know (Theorem 3(a)) that the amount of common information in two strings is not larger than the mutual information of these strings. A natural related question is the following one: can common information be far less than mutual information?

This question was positively answered by Gács and Körner [1]. They found out that there are pairs of strings x and y such that $I(x : y)$ is big but nevertheless any string z that is simple relative to both x and y (both $K(z|x)$ and $K(z|y)$ are small) is simple (has small $K(z)$).

Their construction uses ideas from Shannon information theory. Another construction was suggested in [4] (see [5] for details). Here we present a third way to construct examples of that kind.

Consider a finite field F_n of cardinality $q = q_n$ close to 2^n. (Any field of size $2^{n+O(1)}$ will work, so we may use the field of cardinality 2^n or the field $\mathbb{Z}/q\mathbb{Z}$ where q is a prime number between 2^n and 2^{n+1}.) Consider three-dimensional vector space over F_n. Any non-zero vector (f_1, f_2, f_3) generates a line (by “line” we mean a line going through 0, i.e., one-dimensional subspace). Two lines generated by (f_1, f_2, f_3) and

(g_1, g_2, g_3) are called orthogonal if $f_1g_1 + f_2g_2 + f_3g_3 = 0$. Now consider two random orthogonal lines x and y (i.e. pair of two orthogonal lines (x, y) which has the greatest possible complexity). We claim that $I(x:y)$ is significant but there is no string z which is simple relative to both x and y unless z is simple.

More precisely, consider the set

$$O = \{(x, y)\colon\ x \text{ and } y \text{ are orthogonal lines}\}.$$

This set contains $q^3 + o(q^3)$ elements (there are q^2+q+1 lines and each line is orthogonal to $q+1$ lines). Therefore, O contains a pair (x, y) whose complexity is $\log(q^3(1+o(1))) = 3n + O(1)$. (We assume that elements of F_n are encoded by binary strings of length $n + O(1)$, so we can speak about complexities.) Note that $K(x) \leqslant 2n + O(\log n)$ since there are about 2^{2n} lines; moreover, $K(y|x) \leqslant n + O(\log n)$ since y is one of $2^{n+O(1)}$ lines orthogonal to A. Recalling the inequality $K(\langle x, y\rangle) \leqslant K(x) + K(y|x) + O(\log n)$, we conclude that $K(x) = 2n + O(\log n)$ and $K(y|x) = n + O(\log n)$. For similar reasons $K(y) = 2n + O(\log n)$ and $K(x|y) = n + O(\log n)$. Therefore, $I(x:y) = n + O(\log n)$.

Remark. We would like to caution against free usage of geometrical intuition in our context. For instance, though we use the term "orthogonal", we have no scalar product in linear spaces over finite fields and a nonzero vector may be orthogonal to itself.

Theorem 4. *Let $\langle x_n, y_n\rangle$ be a random pair of orthogonal lines in the three-dimensional space over F_n. For any sequence of strings z_n*

$$K(z_n) \leqslant 2K(z_n|x_n) + 2K(z_n|y_n) + O(\log n)$$

assuming that z_n has polynomial (in n) length. [The constant in $O(\log n)$-notation does not depend on n.]

This theorem implies that sequences $\boldsymbol{x} = x_1, x_2, \ldots$ and $\boldsymbol{y} = y_1, y_2, \ldots$ have $\boldsymbol{\Lambda} = \Lambda, \Lambda, \ldots$ as their greatest lower bound. (Here Λ denotes the empty string.) Indeed, if $K(z_n|x_n) = O(\log n)$ and $K(z_n|y_n) = O(\log n)$ for some sequence $\boldsymbol{z} = z_1, z_2, \ldots$, then $K(z_n) = O(\log n)$ according to Theorem 4.

Proof. Proof of Theorem 4 is based on a simple combinatorial observation.

Lemma 5. *Consider a bipartite graph with k vertices $1, \ldots, k$ on the left and l vertices $1, \ldots, l$ on the right. Assume that for any two different nodes u, v on the left there are at most r nodes on the right connected with both u, v. Then the following bound for the number of edges $|E|$ is valid:*

- $k \leqslant \sqrt{l/r} \Rightarrow |E| \leqslant 2l$;
- $k \geqslant \sqrt{l/r} \Rightarrow |E| \leqslant 2k\sqrt{lr}$.

Indeed, for each element v on the left consider the set N_v of its neighbors on the right; let n_v be the cardinality of N_v. The intersection $N_v \cap N_w$ (for $v \neq w$) contains at most r element. Assume that $k \leqslant \sqrt{l/r}$. Consider the union of all N_v; it has at least

$$n_1 + n_2 + \cdots + n_k - \sum_{i<j} |N_i \cap N_j|$$

elements. On the other hand, it has at most l elements. The number of pairs $\langle i, j \rangle$ is less than $k^2 \leqslant l/r$. Therefore

$$n_1 + n_2 + \cdots + n_k - (l/r)r \leqslant l \Rightarrow |E| = n_1 + n_2 + \cdots + n_k \leqslant 2l.$$

The first statement is proved. It implies that for $k = \sqrt{l/r}$ (we assume here that the number $\sqrt{l/r}$ is integer; the proof can be easily modified to handle the general case) the average number of neighbors for vertices on the left is at most $2\sqrt{lr}$. We use this observation to prove the second part of the lemma.

Let $k \geqslant \sqrt{l/r}$. Consider $\sqrt{l/r}$ vertices on the left having maximum neighborhoods and delete all other vertices on the left; this makes the average number of neighbors bigger. But we know that it does not exceed $2\sqrt{lr}$. The same is true for the initial graph. Therefore $|E| \leqslant k \cdot 2\sqrt{lr}$. Lemma 5 is proved.

This lemma will be applied to a bipartite graph whose vertices (both on the left and on the right) are lines; edges connect pairs of orthogonal lines. It is easy to see that we can let $r = 1$ (if both x, y are orthogonal to both z, u and $x \neq y$ then $z = u$).

Now we are ready to prove Theorem 4. As we know, $K(x) = K(y) = 2n$ and $K(\langle x, y \rangle) = 3n$ (from now on we omit $O(\log n)$-terms for brevity). Let $K(z|x) = p_1$ and $K(z|y) = p_2$. We want to get an upper bound for $m = K(z)$. First, let us compute $K(x|z)$ and $K(y|z)$:

$$K(x|z) = K(\langle x, z \rangle) - K(z) = K(x) + K(z|x) - K(z) = 2n + p_1 - m.$$

Similarly, $K(y|z) = 2n + p_2 - m$. Consider the set P of all lines whose complexity relative to z does not exceed $K(x|z)$; this set contains line x and has cardinality 2^{2n+p_1-m} (up to a polynomial in n factor). Similarly we get a set Q that contains lines whose complexity relative to z does not exceed $K(y|z)$; this set has cardinality 2^{2n+p_2-m}. Consider a bipartite graph whose edges connect orthogonal lines from P and Q. This graph satisfies the lemma for $r = 1$, so the number of edges $|E|$ does not exceed

$$2^{2n+p_2-m} \quad \text{if } (2n + p_1 - m) \leqslant (2n + p_2 - m)/2;$$

$$2^{2n+p_1-m} \cdot \sqrt{2^{2n+p_2-m}} \quad \text{if } (2n + p_1 - m) \geqslant (2n + p_2 - m)/2.$$

On the other hand, the pair $\langle x, y \rangle$ represents one of the edges of that graph. If z is known, we can enumerate P, Q and E, so the pair $\langle x, y \rangle$ may be described by its number in E. Hence $3n = K(\langle x, y \rangle) \leqslant K(z) + \log |E|$. Therefore, the two bounds for $|E|$ imply

$$3n \leqslant m + (2n + p_2 - m) \Rightarrow n \leqslant p_2$$

(the first one) and

$$3n \leqslant m + (2n + p_1 - m) + \tfrac{1}{2}(2n + p_2 - m) \Rightarrow m \leqslant 2p_1 + p_2$$

(the second one). We have to prove that $m \leqslant 2p_1 + 2p_2$ (recall that logarithmic terms are omitted). In the second case it is evident; in the first case one should note that $K(z) \leqslant K(z|x) + K(x) \leqslant p_1 + 2n \leqslant p_1 + 2p_2 \leqslant 2p_1 + 2p_2$. □

Remark. The same example may be reformulated in several ways. Replacing line y by the orthogonal plane $y^{\perp}$, we may say that $\langle x, y\rangle$ is a random pair $\langle$line x, plane y going through $x\rangle$. We may then switch from projective plane to affine plane and say that $\langle x, y\rangle$ is a random pair $\langle$point x on the affine plane, line y that goes through $x\rangle$. Indeed, fix any affine plane P not going through zero. Then x may be identified with the common point of P and x and plane y with the common line of y and P. (We lose lines that are parallel to P, but those lines are not random.) The third way (used in [5]) to reformulate the example is to say that $x = (a, b)$ and $y = (c, ac + b)$ where (a, b, c) is a random triple of elements of F. Indeed, $x = (a, b)$ identifies the affine line $\{(u, v) \mid v = au + b\}$ (again we lose affine lines that are parallel to the line $u = 0$, but all those lines are not random) and $y = (c, ac + b)$ is a point on that line.

Using Lemma 5 we can prove that several other examples of pairs have no common information. Here are two of them:

Theorem 5. (a) *Let $\langle x_n, y_n\rangle$ be a random pair of orthogonal lines in four-dimensional space over F_n. For any sequence of strings z_n*

$$K(z_n) \leqslant 2K(z_n|x_n) + 2K(z_n|y_n) + O(\log n)$$

assuming that z_n has polynomial (*in n*) *length.*

(b) *The same is true if $\langle x_n, y_n\rangle$ is a random pair of intersecting affine lines* (*one-dimensional affine subspaces*) *in three-dimensional affine space over F_n.*

Proof. (a) The proof goes along the same lines as the proof of the previous theorem, so we just outline the main points.

- $K(x) = K(y) = 3n$ and $K(\langle x, y\rangle) = 5n$ (we omit $O(\log n)$-terms). Thus, in this case $K(x|z) = K(x) + K(z|x) - K(z) = 3n + p - m$ and $K(y|z) = 3n + q - m$.
- We consider the same bipartite graph (but now a line means a line in a four-dimensional space). This time the conditions of Lemma 5 are fulfilled for $r = 2^n$, because the number of lines in four-dimensional space orthogonal to two different given lines is 2^n.
- Thus the number of edges $|E|$ does not exceed

$$2^{3n+q-m} \text{ if } (3n + p - m) \leqslant (2n + q - m)/2;$$

$$2^{3n+p-m} \cdot \sqrt{2^{4n+q-m}} \text{ if } (3n + p - m) \geqslant (2n + q - m)/2.$$

- On the other hand, $5n = K(\langle x, y\rangle) \leqslant K(z) + \log |E|$. Therefore, the two bounds for $|E|$ imply

$$5n \leqslant m + (3n + q - m) \Rightarrow 2n \leqslant q$$

(the first one) and

$$5n\leqslant m+(3n+p-m)+\tfrac{1}{2}(4n+q-m)\Rightarrow m\leqslant 2p+q$$

(the second one).

- In the first case one should note that $K(z)\leqslant K(z|x)+K(x)\leqslant p+3n\leqslant p+\frac{3}{2}q\leqslant 2p+2q$.

(b) This time we connect by edges affine lines that have a common point, thus the conditions of the lemma are true for $r=2^{2n}$ (there are this many affine lines intersecting two given different affine lines). The rest is as follows:

- $K(x)=K(y)=4n$ and $K(\langle x,y\rangle)=7n$ (omitting $\mathrm{O}(\log n)$-terms),
- $K(x|z)=4n+p-m$, $K(y|z)=4n+q-m$,
- the number of edges $|E|$ does not exceed 2^{4n+q-m} if $(4n+p-m)\leqslant(2n+q-m)/2$ and $2^{4n+p-m}\cdot\sqrt{2^{6n+q-m}}$ if $(4n+p-m)\geqslant(2n+q-m)/2$,
- hence $7n\leqslant m+(4n+q-m)\Rightarrow 3n\leqslant q$ in the first case and $7n\leqslant m+(4n+p-m)+\frac{1}{2}(6n+q-m)\Rightarrow m\leqslant 2p+q$ in the second case. In the first case one should note that $K(z)\leqslant K(z|x)+K(x)\leqslant p+4n\leqslant p+\frac{4}{3}q\leqslant 2p+2q$. □

Let us note that in these examples some z_n still have more information about x_n and y_n than one could expect. For example, if in (b) we consider the intersection point p_n of x_n and y_n, then $K(p_n)=3n$, $K(x_n|p_n)=2n$, $K(y_n|p_n)=2n$ (omitting $\mathrm{O}(\log n)$-terms). There are some x_n' and y_n' with the same complexities ($K(x_n')=4n$, $K(y_n')=4n$, $K(\langle x_n',y_n'\rangle)=7n$) for which there is no p_n with similar properties. (Remark: Instead of intersection point we could consider two-dimensional affine subspace that contains both lines.)

For (a) one also can find p_n that contain more information about x_n and y_n than one could expect. The way to construct such p_n was pointed by Finkelberg and Bezrukawnikov. Let W be the two-dimensional subspace (a plane) containing the vectors $(1,0,0,0)$ and $(0,1,0,0)$ (the choice of W is not important: any plane W with $K(W)=\mathrm{O}(\log n)$ would work). Let w be any line in W orthogonal to y (obviously it exists). Take as P the plane having the lines x and w (as x is random, $x\notin W$). Let us note that P has 1-dimensional intersection with W and the number of planes with this property is about 2^{3n}, therefore $K(P)\leqslant 3n+\mathrm{O}(\log n)$. The number of lines in P is about 2^n, thus $K(x|P)\leqslant n+\mathrm{O}(\log n)$. The line y is orthogonal to both x,w, therefore this line is orthogonal to P. The number of lines orthogonal to P is about 2^n, therefore $K(y|P)\leqslant n+\mathrm{O}(\log n)$.

This effect (some p contains more information about x and y than one could expect) is analyzed in Section 6.

5. More examples: a new method

The examples of Theorems 4 and 5(a) are specific cases of the following example. Let m,k be integer constants and let x_n and y_n be random orthogonal k-dimensional subspaces of an m-dimensional linear space over F_n. (Recall that F_n denotes a field having about 2^n elements.) If $m<2k$ then there are no orthogonal k-dimensional subspaces. If $m=2k$ then x_n determines y_n uniquely. Hence their greatest lower bound is equal to x_n. So we will assume that $m>2k$. It was proven in [6] that for any such m,k the greatest lower bound of $\boldsymbol{x},\boldsymbol{y}$ is the sequence $\boldsymbol{\Lambda}=\Lambda,\Lambda,\ldots$. Note that the most interesting case is when m is close to $2k$ because then the mutual information of x_n, y_n is close to complexities of both x_n, y_n. Indeed, it is easy to verify that

$$K(x_n)=(mk-k^2)n+\mathrm{O}(\log n),$$

$$K(y_n)=(mk-k^2)n+\mathrm{O}(\log n),$$

$$I(x_n:y_n)=k^2n+\mathrm{O}(\log n).$$

So, the fraction $I(x_n:y_n)/K(x_n)$ is close to 1 as k/m is close to 1/2 (recall that k,m are fixed thus the constants in O-notation may depend on k,m). In this section, we give a new proof of the result of [6] using clearer combinatorial arguments.

Theorem 6 (Romashchenko [6]). *Let $2k<m$ and x_n and y_n be random orthogonal k-dimensional subspaces of an m-dimensional linear space over F_n (where F_n is a field having about 2^n elements). Then there are positive c_1,c_2 such that the following holds. For any sequence of strings z_n such that $K(z_n|x_n)$, $K(z_n|y_n)<c_1n$, we have $K(z_n)\leqslant c_2(K(z_n|x_n)+K(z_n|y_n))+\mathrm{O}(\log n)$. (The constant in* O*-notation may depend on m but not on n.)*

Proof. Recall the proof of Theorem 4. Using a combinatorial property of the graph whose nodes are 1-dimensional subspaces of the 3-dimensional space over F_n and edges connect orthogonal subspaces, we proved that any its subgraph has few edges. (A subgraph of a graph (V,E) is a graph of the form $(U,\ E\cap(U\times U))$ where $U\subseteq V$.) That property stated that any two nodes have at most one common neighbor. Now this property does not hold and we shall define another one. Graphs satisfying that property will be called t,ε-oblivious. (Now we shall consider ordinary undirected graphs, not bipartite ones.) Then we will prove an appropriate analog of Lemma 5 for t,ε-oblivious graphs.

Assume that starting from a node $v\in V$ we make t moves of a random walk in the finite graph (V,E); on every step we move to a random neighbor of the current node. Let $v(t)$ stand for the end node of the walk. The graph is called *t,ε-oblivious* if for any $v\in V$ and for any $U\subseteq V$,

$$\mathrm{Prob}[v(t)\in U]\leqslant\frac{|U|}{|V|}+\varepsilon.$$

Lemma 6. *Let (V,E) be the graph whose nodes are k-dimensional subspaces of the m-dimensional space over F_n and edges connect orthogonal subspaces. Then (V,E)*

is t,ε-oblivious, where $t=2\lceil k/(m-2k)\rceil$, and $\varepsilon=C2^{-n}$ (where C is a positive real depending on m but not on n).

Proof. Let a,b be two subspaces of the m-dimensional space over F_n. It is well known that

$$\dim a + \dim b = \dim(a\cup b) + \dim(a\cap b).$$

Here $a\cup b$ stands for linear sum of a and b. Hence $\dim(a\cap b) \geqslant \dim a + \dim b - m$. Assume that a is fixed, $\dim a = k$, and b is a random l-dimensional subspace. With overwhelming probability the dimension of $a\cap b$ is as low as possible (that is, $\max\{0;\, k+l-m\}$). More precisely, the following claim is true.

Claim 1. *The probability of the event*

$$\dim(a\cap b) = \max\{0; k+l-m\}$$

is at least $(1-C2^n)$ for some positive C depending only on m.

(We postpone the proof of the claim to the end of the proof of the theorem.)

Let a and b be k-dimensional subspaces such that

$$\dim(a\cap a^{\perp}) = r_0,$$

$$\dim(a\cap b) = r_1,$$

$$\dim(a^{\perp}\cap b) = r_2,$$

$$\dim(a\cap a^{\perp}\cap b) = r_3,$$

$$\dim((a\cup a^{\perp})\cap b) = r_4,$$

where $a^{\perp}$ stands for orthogonal complement to a. (Note that intersection of a and $a^{\perp}$ may be nontrivial.) Let c be a random k-dimensional subspace from the orthogonal complement to b.

Claim 2. *For some positive C depending only on m with probability greater than $(1-C2^{-n})$ it holds*

$$\dim(a\cap c) = \max\{0; r_2-(m-2k)\},$$

$$\dim(a^{\perp}\cap c) = r_1,$$

$$\dim(a\cap a^{\perp}\cap c) = \max\{0; r_4-(m-k-r_0)\},$$

$$\dim((a\cup a^{\perp})\cap c) = r_3+k-r_0.$$

Proof. Find first the dimension of intersection of a with the orthogonal complement to b. As $\dim a^{\perp} = m - \dim a = m - k$, we have

$$\dim(a \cap b^{\perp}) = \dim(a^{\perp} \cup b)^{\perp} = m - \dim(a^{\perp} \cup b)$$

$$= m - (\dim a^{\perp} + \dim b - \dim(a^{\perp} \cap b)) = m - ((m-k) + k - r_2) = r_2.$$

As $a \cap c = (a \cap b^{\perp}) \cap c$ we can find the most probable dimension of $a \cap c$ by applying Claim 1 to subspaces $a \cap b^{\perp}$ and c of the linear space $b^{\perp}$. Thus we obtain

$$\dim(a \cap c) = \max\{0; r_2 + k - (m-k)\} = \max\{0; r_2 - (m - 2k)\}$$

with probability at least $(1 - C2^{-n})$. In a similar way we find the most probable dimension of intersection of subspaces $a^{\perp}$ and c. We have

$$\dim(a^{\perp} \cap b^{\perp}) = \dim(a \cup b)^{\perp} = m - \dim(a \cup b)$$

$$= m - (\dim a + \dim b - \dim(a \cap b)) = m - (k + k - r_1) = m - 2k + r_1.$$

Applying Claim 1 to subspaces $a^{\perp} \cap b^{\perp}$ and c of linear space $b^{\perp}$ we see that

$$\dim(a^{\perp} \cap c) = \max\{0; (m - 2k + r_1) + k - (m-k)\} = \max\{0; r_1\}$$

with probability at least $(1 - C2^{-n})$. In a similar way we obtain

$$\dim(a \cap a^{\perp} \cap b^{\perp}) = m - \dim(a \cup a^{\perp} \cup b)$$

$$= m - \dim(a \cup a^{\perp}) - \dim b + \dim((a \cup a^{\perp}) \cap b)$$

$$= m - (m - r_0) - k + r_4 = r_0 - k + r_4.$$

Thus

$$\dim(a \cap a^{\perp} \cap c) = \max\{0; (r_0 - k + r_4) + k - (m-k)\}$$

$$= \max\{0; k + r_4 - m + r_0\}$$

with probability at least $(1 - C2^{-n})$. Finally,

$$\dim((a \cup a^{\perp}) \cap b^{\perp}) = m - \dim((a \cap a^{\perp}) \cup b)$$

$$= m - \dim(a \cap a^{\perp}) - \dim b + \dim(a \cap a^{\perp} \cap b)$$

$$= m - r_0 - k + r_3.$$

Thus

$$\dim((a \cup a^{\perp}) \cap c) = \max\{0; (m - r_0 - k + r_3) + k - (m-k)\} = r_3 + k - r_0$$

with probability at least $(1 - C2^{-n})$. The claim is proven.

Fix an arbitrary $v \in V$. Denote by r_0 dimension of intersection $v \cap v^{\perp}$. Let S_i stand for the set of all $u \in V$ such that

$$\dim(v \cap u) = r_1(i),$$

$$\dim(v^{\perp} \cap u) = r_2(i),$$

$$\dim(v \cap v^{\perp} \cap u) = r_3(i),$$

$$\dim((v \cup v^{\perp}) \cap u) = r_4(i),$$

where $r_1(0) = k$, $r_2(0) = r_0$, $r_3(0) = r_0$, $r_4(0) = k$, and

$$r_1(i+1) = \max\{0; r_2(i) - (m - 2k)\},$$

$$r_2(i+1) = r_1(i),$$

$$r_3(i+1) = \max\{0; r_4(i) - (m - k - r_0)\},$$

$$r_4(i+1) = r_3(i) + k - r_0.$$

The above recurrence implies that

$$r_1(i+2) = \max\{0; r_1(i) - (m - 2k)\},$$

$$r_2(i+2) = \max\{0; r_2(i) - (m - 2k)\},$$

$$r_3(i+2) = \max\{0; r_3(i) - (m - 2k)\},$$

$$r_4(i+2) = \max\{k - r_0; r_4(i) - (m - 2k)\}.$$

Hence $r_1(t) = r_2(t) = r_3(t) = 0$, $r_4(t) = k - r_0$ (recall that $t = 2\lceil k/(m-2k) \rceil$). By Claim 1 the probability for a random $x \in V$ to get into S_t is at least $1 - C2^{-n}$ for some C depending only on m.

Let $v(i)$, $i \leqslant t$ denote the ith node in a random walk starting from v (and $v(0) = v$). Let G_i stand for the event

$$v(0) \in S_0, v(1) \in S_1, \ldots, v(i) \in S_i.$$

Using Claim 2 it is easy to prove by induction that for any $v \in V$ the probability of G_i is at least $1 - C2^{-n}$ (where C depends on m only).

Claim 3. *Let a, b and c be as in Claim 2. The probability of event*

$$\dim(a \cap c) = q_1,$$

$$\dim(a^{\perp} \cap c) = q_2,$$

$$\dim(a \cap a^{\perp} \cap c) = q_3,$$

$$\dim((a \cup a^{\perp}) \cap c) = q_4,$$

is a function of $k, r_0, r_1, r_2, r_3, r_4, q_1, q_2, q_3, q_4$ (*but it does not depend on the choice of* a *and* b).

(We postpone the proof of the claim to the end of the proof of the theorem.)

Claim 4. *The probability* $\mathrm{Prob}[v(i) = u_i | G_i]$ *is the same for all* $u_i \in S_i$ (*and hence is equal to* $1/|S_i|$).

Proof. The proof is by induction on i. For $i = 0$ the statement is trivial. Let $i > 0$ and $u_i \in S_i$. We have

$$\mathrm{Prob}[v(i) = u_i | G_i] = \mathrm{Prob}[v(i) = u_i | G_{i-1}] \frac{\mathrm{Prob}[G_{i-1}]}{\mathrm{Prob}[G_i]}.$$

The second factor does not depend on u_i, so it remains to prove that neither does the first factor. Let $U_i^{\perp}$ denote the set of all $u \in V$ orthogonal to u_i. We have

$$\mathrm{Prob}[v(i) = u_i | G_{i-1}] = \sum_{u_{i-1} \in S_{i-1} \cap U_i^{\perp}} \frac{\mathrm{Prob}[v(i-1) = u_{i-1} | G_{i-1}]}{M},$$

where M stands for the number of k-dimensional subspaces orthogonal to a fixed k-dimensional subspace. By induction hypothesis the numerator of the last fraction is equal to $1/|S_{i-1}|$. Therefore we have

$$\mathrm{Prob}[v(i) = u_i | G_{i-1}] = \sum_{u_{i-1} \in S_{i-1} \cap U_i^{\perp}} \frac{1}{M|S_{i-1}|} = \frac{|S_{i-1} \cap U_i^{\perp}|}{M|S_{i-1}|}.$$

The factor $|S_{i-1} \cap U_i^{\perp}|/M$ is equal to the probability of the event "a random $x \in V$ orthogonal to u_i belongs to S_{i-1}". By Claim 3 this probability does not depend on $u_i \in S_i$. Claim 4 is now proved.

By Claim 4 for any $U \subseteq V$ we have

$$\mathrm{Prob}[v(t) \in U | G_t] = |U \cap S_t| / |S_t|$$

Therefore,

$$\begin{aligned} \mathrm{Prob}[v(t) \in U] &= \mathrm{Prob}[v(t) \in U | G_t] \cdot \mathrm{Prob}[G_t] + \mathrm{Prob}[v(t) \in U, \bar{G}_t] \\ &\leqslant |U \cap S_t| / |S_t| + \mathrm{Prob}[\bar{G}_t]. \end{aligned}$$

The second term is bounded by $C2^{-n}$. Estimate the first term:

$$\frac{|U \cap S_t|}{|S_t|} \leqslant \frac{|U|}{|V|(1 - C2^{-n})} \leqslant \frac{|U|}{|V|} + C'2^{-n}. \qquad \square$$

Lemma 7. *Assume that every node in a t,ε-oblivious graph (V,E) has degree d or less. Then the number of edges in any subgraph U of V is at most $\alpha|U|$, where*

$$\alpha = \left\lfloor d\left(\frac{|U|}{|V|}+\varepsilon\right)^{1/t}\right\rfloor.$$

Proof. Define $U' \subseteq U$ as follows. Let us start with $U' = \emptyset$ and iterate the following step: if there is a node $v \in U \backslash U'$ that has at most α adjacent nodes in $U \backslash U'$ then choose any such node and include it in U'. Otherwise halt. The resulting subgraph U' has at most $\alpha|U'|$ edges, as on each step the number of edges that are incident to some node in U' increases at most by α. Another useful property of U' is as follows: any node $v \in U\backslash U'$ has at least $\alpha+1$ neighbors in the set $U\backslash U'$. Let us prove that actually U' coincides with U. Suppose this is not true. Then choose a node $v \in U\backslash U'$. We have

$$\text{Prob}[v(t) \in U\backslash U'] \geqslant \left(\frac{\alpha+1}{d}\right)^t > \frac{|U|}{|V|}+\varepsilon.$$

On the other hand,

$$\text{Prob}[v(t) \in U\backslash U'] \leqslant \text{Prob}[v(t) \in U] \leqslant \frac{|U|}{|V|}+\varepsilon.$$

These two inequalities are inconsistent, this proves that $U' = U$. Thus the number of edges in U is at most $\alpha|U|$. □

Lemma 8. *Let t be an integer number and $0<\varepsilon<1$ a real number. Let $G=(V,E)$ be a t,ε-oblivious graph in which any node has degree d. Let (u,v) be a random edge in G (that is, $K(u,v|G) \geqslant \log|E|$) and let z be a string. Then at least one of the following three inequalities holds*:

$$K(z|u,G) \geqslant \frac{1}{t}\left(\log\frac{1}{\varepsilon}-1\right),$$

$$K(z|v,G) \geqslant \frac{1}{t}\left(\log\frac{1}{\varepsilon}-1\right),$$

$$K(z|G) < (t+1)\max\{K(z|u,G), K(z|v,G)\} + \mathrm{O}\left(\log\left(\log\frac{1}{\varepsilon}+\log|V|\right)\right)$$

(*the constant in* O-*notation does not depend on* t,ε).

Proof. Assume that the first two inequalities are false. Let

$$k = \max\{K(z|u,G), K(z|v,G)\}, \quad m = K(z|G).$$

We have $\varepsilon < 2^{-kt-1}$. First estimate m very roughly:

$$m = K(z|G) \leqslant K(z|u,G) + 2K(u|G) + \mathrm{O}(1) \leqslant \frac{1}{t}\log\frac{1}{\varepsilon} + 2\log|V| + \mathrm{O}(1).$$

Thus the complexities of all u, v, z conditional to G are polynomial in $\log|V|, \log\frac{1}{\varepsilon}$. In what follows we omit additive $O(\log(\log|V| + \log\frac{1}{\varepsilon}))$ terms. We have

$$K(u|z,G) = K(u|G) + K(z|u,G) - K(z|G) \leqslant \log|V| + k - m.$$

The same bound is valid for $K(v|z,G)$.

Let U be the set of all $x \in V$ such that $K(x|z,G) \leqslant K(u|z,G)$, $K(v|z,G)$. Then $|U| \leqslant |V|2^{k-m}$ (up to a factor polynomial in $\log|V|, \log\frac{1}{\varepsilon}$). By Lemma 7 we obtain the following upper bound for the number E_U of edges in U:

$$|E_U| \leqslant d|U| \left(\frac{|U|}{|V|} + \varepsilon\right)^{1/t}.$$

As $u, v \in U$ and U (hence E_U) is enumerable given $z, G, K(u|z,G), K(v|z,G)$, we have

$$\begin{aligned}\log(|V|d/2) = \log|E| \leqslant K(u,v|G) &\leqslant \log|E_U| + K(z|G)\\ &\leqslant \log d + \log|U| + \log\left(\frac{|U|}{|V|} + \varepsilon\right)^{1/t} + m\\ &\leqslant \log d + \log|V| + (k-m) + (1/t)\log(2^{k-m} + \varepsilon) + m.\end{aligned}$$

Therefore, we have

$$2^{-kt} \leqslant 2^{k-m} + \varepsilon$$

(up to a factor polynomial in $\log|V|$, $\log(1/\varepsilon)$). By our assumption ε is less than half of 2^{-kt}. Hence

$$-kt \leqslant k - m \Rightarrow m \leqslant (t+1)k. \qquad \square$$

The assertion of the theorem is a direct corollary of the proven lemmas.

Thus it remains to prove Claims 1 and 3.

Proof of Claim 1. Let N stand for the number of elements in the field F_n (recall that $N \approx 2^n$). Let u be an i-dimensional subspace of the m-dimensional space over F_n. The number of vectors that do not belong to u is equal to $N^m - N^i = N^m(1 + O(1/N))$ (provided $i < m$). Assume that $i + l \leqslant m$. The number Seq_l^{mi} of sequences of vectors $e_1, \dots, e_l$ such that the system

$$(\text{a basis of } u) \cup \{e_1, \dots, e_l\}$$

is independent is equal to $N^{ml}(1 + O(1/N))$ (the constant in O-notation depends on l).

Let Sub_l^m stand for the number of l-dimensional subspaces of the m-dimensional space. We have

$$Sub_l^m = \frac{Seq_l^{m0}}{Seq_l^{l0}} = \frac{N^{ml}(1 + O(1/N))}{N^{l^2}(1 + O(1/N))} = N^{(m-l)l}(1 + O(1/N)).$$

Let a be a k-dimensional subspace. The number of l-dimensional subspaces b such that $\dim(a\cap b)=s$ is equal to the number of s-dimensional subspaces c of a multiplied by the number of l-dimensional subspaces b whose intersection with a is equal to a fixed s-dimensional subspace c:

$$Sub_s^k \frac{Seq_{l-s}^{mk}}{Seq_{l-s}^{ls}}.$$

Hence the probability that a random l-dimensional subspace b satisfies the equality $\dim(a\cap b)=s$ is equal to

$$\frac{Sub_s^k Seq_{l-s}^{mk}}{Seq_{l-s}^{ls} Sub_l^m} = \frac{N^{(k-s)s}N^{m(l-s)}}{N^{l(l-s)}N^{(m-l)l}}(1+\mathrm{O}(1/N))$$

$$= N^{(k+l-m-s)s}(1+\mathrm{O}(1/N)).$$

This probability is exponentially (in n) close to 1 when either $s=0$ or $s=k+l-m$.

Proof of Claim 3. Assume that

$$\dim(a\cap b)=r_1,\ \dim(a^{\perp}\cap b)=r_2,\ \dim(a\cap a^{\perp}\cap b)=r_3,$$

$$\dim((a\cup a^{\perp})\cap b)=r_4.$$

Then

$$\dim(a\cap b^{\perp})=r_2,$$

$$\dim(a^{\perp}\cap b^{\perp})=m-2k+r_1,$$

$$\dim(a\cap a^{\perp}\cap b^{\perp})=r_0-k+r_4,$$

$$\dim((a\cup a^{\perp})\cap b^{\perp})=m-r_0-k+r_3.$$

Thus the claim is a particular case of the following general fact.

> Let α,β,γ be subspaces of a linear space L over a finite field F such that $\alpha\cup\beta\subseteq\gamma$. Then the probability for a random k-dimensional subspace δ of L of satisfying the equalities $\dim(\delta\cap\alpha)=q_1$, $\dim(\delta\cap\beta)=q_2$, $\dim(\delta\cap\alpha\cap\beta)=q_3$, $\dim(\delta\cap\gamma)=q_4$, depends only on $k,q_1,q_2,q_3,q_4,\dim\alpha,\dim\beta,\dim(\alpha\cap\beta),\dim\gamma,\dim L,|F|$.

(We apply this assertion to $\alpha=a\cap b^{\perp}$, $\beta=a^{\perp}\cap b^{\perp}$, $\gamma=(a\cup a^{\perp})\cap b^{\perp}$, $L=b^{\perp}$.)

Proof. Let α',β',γ' be a triple of linear subspaces such that $\alpha'\cup\beta'\subseteq\gamma'$ and $\dim\alpha'=\dim\alpha$, $\dim\beta'=\dim\beta$, $\dim(\alpha'\cap\beta')=\dim(\alpha\cap\beta)$, $\dim\gamma'=\dim\gamma$. Then there is an automorphism φ of L such that $\varphi\alpha=\alpha'$, $\varphi\beta=\beta'$, $\varphi\gamma=\gamma'$. Indeed, construct five systems of vectors $A_1,A_2,\dots,A_5$ as follows. The first system, A_1 is a basis of $\alpha\cap\beta$. The second system, A_2 completes A_1 to the basis of α. The third system, A_3 completes A_1 to the basis of β. It is easy to see that $A_1 \cup A_2 \cup A_3$ is a basis of $\alpha \cup \beta$. The fourth system, A_4 completes this union to the basis of γ. The fifth system, A_5 completes the union of the four defined systems to the basis of L. In a similar way construct five systems $A'_1, A'_2, \dots, A'_5$ for α', β', γ'. The assumptions on dimensions of subspaces guarantee that A_i and A'_i have the same number of elements. The automorphism φ is generated by one to one correspondence between A_i and A'_i.

Thus we have

$$\begin{aligned}
&\mathrm{Prob}[\dim(\delta \cap \alpha) = q_1,\ \dim(\delta \cap \beta) = q_2,\\
&\quad \dim(\delta \cap \alpha \cap \beta) = q_3,\ \dim(\delta \cap \gamma) = q_4]\\
&= \mathrm{Prob}[\dim \varphi(\delta \cap \alpha) = q_1,\ \dim \varphi(\delta \cap \beta) = q_2,\\
&\quad \dim \varphi(\delta \cap \alpha \cap \beta) = q_3,\ \dim \varphi(\delta \cap \gamma) = q_4]\\
&= \mathrm{Prob}[\dim(\varphi\delta \cap \varphi\alpha) = q_1,\ \dim(\varphi\delta \cap \varphi\beta) = q_2,\\
&\quad \dim(\varphi\delta \cap \varphi\alpha \cap \varphi\beta) = q_3,\ \dim(\varphi\delta \cap \varphi\gamma) = q_4]\\
&= \mathrm{Prob}[\dim(\delta \cap \alpha') = q_1,\ \dim(\delta \cap \beta') = q_2,\\
&\quad \dim(\delta \cap \alpha' \cap \beta') = q_3,\ \dim(\delta \cap \gamma') = q_4]. \qquad \square
\end{aligned}$$

6. More about common information

Let us reformulate our informal definition of common information. We say that strings x and y have u-bit common information z if $K(z) \leqslant u$, $K(x|z) \leqslant K(x) - u$, and $K(y|z) \leqslant K(y) - u$. (It is easy to see that all three inequalities in fact are equalities in that case.)

The question whether such z exists is a special case of a more general question: we may ask for given u, v, w whether there is a string z such that $K(z) \leqslant u$, $K(x|z) \leqslant v$, and $K(y|z) \leqslant w$. The set of all triples (u, v, w) for which such a z exists could be considered as "complexity profile" of the pair x, y.

Technically speaking, we should consider sequences of strings instead of individual strings. Let $\boldsymbol{x} = x_1, x_2, \dots$ and $\boldsymbol{y} = y_1, y_2, \dots$ be two sequences such that $|x_n| = O(n)$ and $|y_n| = O(n)$. (Only sequences satisfying these conditions will be considered in this section.) A triple of reals (u, v, w) is called *$\boldsymbol{x}, \boldsymbol{y}$-admissible*, if there exists a sequence $\boldsymbol{z} = z_1, z_2, \dots$ and a constant c such that

$$K(z_n) \leqslant un + c \log n,\ K(x_n|z_n) \leqslant vn + c\ \log n,\ K(y_n|z_n) \leqslant wn + c\ \log n \tag{6}$$

for all n. A triple of reals (u, v, w) is called *$\boldsymbol{x}, \boldsymbol{y}$-non-admissible*, if for any c and for all sufficiently large n there is no z_n satisfying (6) (we consider triples of non-negative reals only). Note that no triple can be $\boldsymbol{x}, \boldsymbol{y}$-admissible and $\boldsymbol{x}, \boldsymbol{y}$-non-admissible simultaneously. But it may happen that a triple falls in neither of these two categories (below we shall give such an example).

The set of all $\boldsymbol{x},\boldsymbol{y}$-admissible triples is denoted by $M^+_{\boldsymbol{x},\boldsymbol{y}}$. The larger $M^+_{\boldsymbol{x},\boldsymbol{y}}$ is, the more information $\boldsymbol{x}$ and $\boldsymbol{y}$ share. The set of all $\boldsymbol{x},\boldsymbol{y}$-non-admissible triples is denoted by $M^-_{\boldsymbol{x},\boldsymbol{y}}$.

Here is a trivial example: assume that x_n is a random string of length n and $y_n = x_n$. Then

$$M^+_{\boldsymbol{x},\boldsymbol{y}} = \{(u,v,w) \mid u+v \geqslant 1,\ u+w \geqslant 1\}, \quad M^-_{\boldsymbol{x},\boldsymbol{y}} = [0,\infty)^3 \setminus M^+_{\boldsymbol{x},\boldsymbol{y}}.$$

If x_n, y_n are random independent strings of length n, then $M^+_{\boldsymbol{x},\boldsymbol{y}}$ is much smaller:

$$M^+_{\boldsymbol{x},\boldsymbol{y}} = \{(u,v,w) \mid u+v \geqslant 1,\ u+w \geqslant 1,\ u+v+w \geqslant 2\},$$

$$M^-_{\boldsymbol{x},\boldsymbol{y}} = [0,\infty)^3 \setminus M^+_{\boldsymbol{x},\boldsymbol{y}}.$$

If x_n, y_n are random strings of length n such that $x_n = y_n$ for even n and x_n, y_n are independent for odd n then

$$M^+_{\boldsymbol{x},\boldsymbol{y}} = \{(u,v,w) \mid u+v \geqslant 1,\ u+w \geqslant 1,\ u+v+w \geqslant 2\},$$

$$M^-_{\boldsymbol{x},\boldsymbol{y}} = \{(u,v,w) \mid u+v < 1 \text{ or } u+w < 1\}$$

(so in this example $M^+_{\boldsymbol{x},\boldsymbol{y}}$ and $M^-_{\boldsymbol{x},\boldsymbol{y}}$ are not complementary). As we shall see, the values of $K(x_n)$, $K(y_n)$ and $K(\langle x_n, y_n\rangle)$ do not determine the sets $M^+_{\boldsymbol{x},\boldsymbol{y}}$, $M^-_{\boldsymbol{x},\boldsymbol{y}}$ completely.

For simplicity we restrict ourselves to one special case: we assume that

$$K(x_n) = 2n + O(\log n),\ K(y_n) = 2n + O(\log n),$$

$$K(\langle x_n, y_n\rangle) = 3n + O(\log n). \tag{7}$$

Consider the following two sets of triples. The first one, called $M^-_{\min}$, contains all the triples satisfying *at least one* of the inequalities

$$u+v+w < 3,\ u+v < 2,\ u+w < 2. \tag{8}$$

The second one, called $M^+_{\min}$, contains all the triples outside $M^-_{\min}$ satisfying *at least one* of the inequalities

$$u+v+w \geqslant 4,\ u+v \geqslant 3,\ u+w \geqslant 3. \tag{9}$$

Theorem 7. (a) *For any sequences $\boldsymbol{x},\boldsymbol{y}$ satisfying* (7)

$$M^+_{\min} \subseteq M^+_{\boldsymbol{x},\boldsymbol{y}}, \qquad M^-_{\min} \subseteq M^-_{\boldsymbol{x},\boldsymbol{y}}.$$

(b) *There exist sequences $\boldsymbol{x},\boldsymbol{y}$ satisfying* (7) *such that $M^+_{\boldsymbol{x},\boldsymbol{y}} = [0,\infty)^3 \setminus M^-_{\min}$ (hence $M^-_{\boldsymbol{x},\boldsymbol{y}} = M^-_{\min}$).*

(c) *There exist sequences* $\boldsymbol{x}, \boldsymbol{y}$ *satisfying* (7) *such that* $M^-_{\boldsymbol{x},\boldsymbol{y}} = [0, \infty)^3 \setminus M^+_{\min}$ (*hence* $M^+_{\boldsymbol{x},\boldsymbol{y}} = M^+_{\min}$).

Proof. (a) Using the inequalities

$$K(\langle x_n, y_n\rangle) \leqslant K(z_n) + K(x_n|z_n) + K(y_n|z_n) + \mathrm{O}(\log n)$$

and $K(x_n) \leqslant K(z_n) + K(x_n|z_n) + \mathrm{O}(\log n)$ we see that inequalities (6) and (7) imply

$$3n + \mathrm{O}(\log n) \leqslant un + vn + wn + \mathrm{O}(\log n),$$

$$2n + \mathrm{O}(\log n) \leqslant un + vn + \mathrm{O}(\log n),$$

$$2n + \mathrm{O}(\log n) \leqslant un + wn + \mathrm{O}(\log n).$$

Hence if at least one of the inequalities

$$3 \leqslant u + v + w, \ 2 \leqslant u + v, \ 2 \leqslant u + w \tag{10}$$

is not fulfilled the triple (u, v, w) is $\boldsymbol{x},\boldsymbol{y}$-non-admissible. Thus, for every $\boldsymbol{x},\boldsymbol{y}$ the set $M^-_{\boldsymbol{x},\boldsymbol{y}}$ includes the set $M^-_{\min}$.

Let us prove that $M^+_{\min} \subseteq M_{\boldsymbol{x},\boldsymbol{y}}$. Without loss of generality assume that $|x_n| = 2n + \mathrm{O}(\log n)$, $|y_n| = 2n + \mathrm{O}(\log n)$ (otherwise replace x_n and y_n by minimum length programs to compute them). Let (u, v, w) be in $M^+_{\min}$. Then the triple (u, v, w) satisfies all the inequalities (10) and at least one of the inequalities (9). So consider three cases.

(1) $u + v + w \geqslant 4$: If $v, w \leqslant 2$ let z be the concatenation of the first $2n - vn$ bits of x and the first $2n - wn$ bits of y (we omit logarithmic terms). Since $u + v + w \geqslant 4$, we have $|z| = 2n - vn + 2n - wn \leqslant un$. To obtain x given z we need the remaining vn bits of x and the numbers $n, \lceil vn \rceil, \lceil wn \rceil$, so $K(x|z) \leqslant vn$. Analogously, $K(y|z) \leqslant wn$.

Otherwise, if say $v > 2$, let z consist of the first un bits of y (and $z = y$ if $2 < u$). Then $K(y|z) \leqslant 2n - un \leqslant wn$ (if $u \leqslant 2$, and $K(y|z) = 0 \leqslant wn$ otherwise). And $K(x|z) \leqslant K(x) \leqslant 2n \leqslant vn$.

(2) $u + v \geqslant 3$: If $u \leqslant 2$ let z consist of the first un bits of y. To find x given z it suffices to know the remaining $2n - un$ bits of y and the minimum program to compute x given y (having n bits). So to find x given z it suffices to have $2n - un + n \leqslant vn$ extra bits. And $K(y|z) \leqslant 2n - un \leqslant wn$.

Otherwise (if $u > 2$) let z be the concatenation of y and the first $un - 2n$ bits of minimum length program p to compute x given y (and $z = yp$ if $un - 2n > n$). To obtain x given z it suffices to have the remaining $n - (un - 2n) \leqslant vn$ bits of p.

(3) $u + w \geqslant 3$: Similar to the previous case.

(b) Let $x_n = pq$, $y_n = pr$, where p, q, r are random independent strings of length n. We have to prove that any triple satisfying the inequalities (10) is $\boldsymbol{x},\boldsymbol{y}$-admissible.

If $u \leqslant 1$ let z consist of the first un bits of p. To find x [y] given z it suffices to have the remaining $n - un$ bits of p and the whole string q [r]. So the total number of bits is $n - un + n \leqslant vn$ [$n - un + n \leqslant wn$].

If $u > 1$ and $v \geqslant 1$ let z consist of the first un bits of y. To find x given z it suffices to have q (n bits). To find y given z it suffices to have the remaining $2n - un$ bits of y and $2n - un \leqslant wn$.

If $u > 1$ and $w \geqslant 1$ use the same argument.

If $u > 1$ and $v, w < 1$ let z be the concatenation of p, the first $n - vn$ bits of q and the first $n - wn$ bits of r. The length of z is $n + n - vn + n - wn \leqslant un$. To find x [y] given z it suffices to have the remaining vn [wn] bits of q [r].

The proven fact agrees with our intuition that these $\mathbf{x}$ and $\mathbf{y}$ have as much common information as possible (under restriction (7)).

(c) This is the most interesting part of the theorem; the proof uses methods from [5].

The set $[0,\infty)^3 \setminus M^+_{\min}$ consists $M^-_{\min}$ and of those triples satisfying the inequalities

$$u + v + w < 4, \; u + v < 3, \; u + w < 3. \tag{11}$$

By item (a) we have $M^-_{\min} \subseteq M^-_{\mathbf{x},\mathbf{y}}$. Therefore, it suffices to prove that any triple satisfying (11) belongs to $M^-_{\mathbf{x},\mathbf{y}}$. Let (u, v, w) satisfy (11). Note that all three inequalities are strict. Assume that for infinitely many n there is z_n for which inequalities (6) are true. Then for infinitely many n,

$$K(z_n) + K(x_n|z_n) + K(y_n|z_n) < 4n, \tag{12}$$

$$K(z_n) + K(x_n|z_n) < 3n, \tag{13}$$

$$K(z_n) + K(y_n|z_n) < 3n. \tag{14}$$

Therefore, it suffices to prove the following lemma.

Lemma 9. *There are* $\mathbf{x}, \mathbf{y}$ *satisfying* (7) *such that for all but finitely many* n *there is no* z_n *satisfying inequalities* (12)–(14).

Proof. Let us fix a natural number n. As usually we will omit the subscript n in x_n, y_n, etc.

We choose the pair (x, y) from the set U consisting of pairs of strings of length $2n + 2\log n$. So $|U| = 2^{4n}n^4$. First remove from U all pairs satisfying at least one of the following requirements:

- $K(x) < 2n$,
- $K(y) < 2n$,
- $K(\langle x, y\rangle) < 3n$,
- there is z satisfying inequalities (12)–(14).

Let us count the number of pairs removed from U to show that U does not become empty. Indeed, less than $2^{2n}2^{2n}n^2$ pairs have been removed for the first reason (and the same amount for the second one), less than 2^{3n} for the third reason and less than $(4n)^3 2^{4n}$ for the fourth reason (for any k, l, m there are at most $2^k 2^l 2^m$ pairs x, y such that there is z with $K(z) = k$, $K(x|z) = l$, $K(y|z) = m$; and the number of triples k, l, m

satisfying the inequality $k + l + m < 4n$ is less than $(4n)^3$). Thus the total number of removed pairs is less than

$$2 \times 2^{4n}n^2 + 2^{3n} + (4n)^3 2^{4n} < 2^{4n}n^4$$

(for sufficiently large n).

Let (x, y) be the least pair remaining in U (with respect to any fixed well founded order). Then $K(x) = 2n + O(\log n)$, $K(y) = 2n + O(\log n)$, $K(\langle x, y\rangle) \geq 3n$ and there is no z satisfying inequalities (12)–(14). Thus, to prove the lemma it suffices to show that $K(\langle x, y\rangle) \leq 3n + O(\log n)$.

Let $W_{k,l}$ stand for the set consisting of all pairs (a, b) such that $K(a) \leqslant k$ and $K(b|a) \leqslant l$. To identify (x, y) it suffices to know n, the set $\{x' | K(x') < 2n\}$, the set $\{(x', y') | K(\langle x', y'\rangle) < 3n\}$, and the sets $W_{k,l}$ for all $k + l < 3n$. The elements of these sets can be enumerated given n. Therefore to get the lists of all these sets it suffices to know n and the number

$$m = |\{x' \mid K(x') < 2n\}| + |\{(x', y') \mid K(\langle x', y'\rangle) < 3n\}| + \sum_{k+l<3n} |W_{k,l}|$$

(given n we enumerate all these sets until m elements have been enumerated; if a pair belongs to several sets we count it separately for each set). As

$$m < 2^{2n} + 2^{3n} + \sum_{k+l<3n} 2^{k+l+2} < 2^{3n+1} + \sum_{j<3n} (j+1)2^{j+2} \leqslant (3n)^2 2^{3n+2},$$

we get

$$K(\langle x, y\rangle) \leqslant \log m + O(\log n) \leqslant 3n + O(\log n). \qquad \square$$

The proof of Theorem 7(c) is non-constructive, it gives no “example” of the pair $(\boldsymbol{x}, \boldsymbol{y})$ with $M^-_{\boldsymbol{x},\boldsymbol{y}} = [0, \infty)^3 \setminus M^+_{\min}$. An example would be a computable sequence of finite non-empty sets A_n of low complexity (say $O(\log n)$) such that any random pair (x_n, y_n) in A_n satisfies Theorem 7(c). Such an example was recently constructed by An. A. Muchnik (unpublished).

In Section 4 we presented several examples of sequences $\boldsymbol{x}, \boldsymbol{y}$ whose common information is less than mutual information. It would be interesting to find the complexity profile for these examples. Unfortunately, we know only few things. We present here known facts about random orthogonal lines in three-dimensional space. In the rest of the paper let $\boldsymbol{x}, \boldsymbol{y}$ be sequences mentioned in Theorem 4. Using Lemma 5 we obtain the following lower bound for $M^-_{\boldsymbol{x},\boldsymbol{y}}$.

Theorem 8. *The set* $M^-_{\boldsymbol{x},\boldsymbol{y}}$ *contains any triple* (u, v, w) *such that* $u + v/2 + \max\{w, v/2\} < 3$ *or* $u + w/2 + \max\{v, w/2\} < 3$.

Note that there are such triples outside $M^-_{\min}$ (for instance, the triple $(1.1, 1.1, 1.1)$).

Proof. Assume that $u + w/2 + \max\{w/2, v\} < 3$ (the other case is entirely similar). Assume that for some c for infinitely many n there is z_n such that (6) holds. Fix any such n (in the sequel we omit subscript n in x_n, y_n, z_n). We use Lemma 5 for the same bipartite graph as in Theorem 4: left nodes are lines having complexity at most $K(x|z)$ conditional to z, right nodes are lines having complexity at most $K(y|z)$ conditional to z, and edges connect orthogonal lines. By Lemma 5 we have

$$|E| \leqslant 2\sqrt{l}\,\max\{\sqrt{l}, k\}.$$

Therefore (we omit logarithmic terms)

$$3n \leqslant K(\langle x, y\rangle) \leqslant \log|E| + K(z) \leqslant (1/2)\log l + \log\max\{\sqrt{l}, k\} + K(z)$$

$$\leqslant wn/2 + \max\{w/2, v\}n + un.$$

As this is true for infinitely many n (up to $\mathrm{O}(\log n)$ term) we get $3 \leqslant u + w/2 + \max\{w/2, v\}$, a contradiction. □

A natural question is whether the inclusion $M^+_{\min} \subseteq M^+_{\boldsymbol{x},\boldsymbol{y}}$ is also strict. The answer to this question may depend on the choice of the field F_n. Note that all proven theorems on $\boldsymbol{x},\boldsymbol{y}$ are true for any choice of F_n. However, it turns out that if the field F_n has size p^2, where p is an integer then the set $M^+_{\boldsymbol{x},\boldsymbol{y}}$ contains the triple $(1.5, 1, 1)$ that is outside $M^+_{\min}$. But we do not know whether this is true for other fields. Recall that we gave similar examples for $\boldsymbol{x},\boldsymbol{y}$ from Theorem 5. To obtain such examples we used arguments from linear algebra. Such arguments can provide only triples (u, v, w) whose coefficients are dimensions of certain linear spaces. But the first component of the triple $(1.5, 1, 1)$ is not an integer so now we cannot use linear-algebraic arguments directly. To overcome this difficulty we "double" the dimension by regarding three-dimensional linear space over F_n as a six-dimensional linear space over the subfield G_n of F_n of size p. Now 1.5 may be obtained as dimension 3 of a space over G_n.

Theorem 9. *Assume that all fields F_n are of size p_n^2 where p_n are integers. Then $M^+_{\boldsymbol{x},\boldsymbol{y}}$ contains the triple* $(1.5, 1, 1)$.

Note that together with previous theorem this implies that in the case $|F_n| = p_n^2$ the triple $(1.5, 1, 1)$ is on the border line between $M^-_{\boldsymbol{x},\boldsymbol{y}}$ and $M^+_{\boldsymbol{x},\boldsymbol{y}}$.

Proof. We shall use the following representation of the example x, y of Theorem 4: $x = (a, b)$, $y = (c, ac + b)$ where (a, b, c) is a random triple of elements of F_n. The pair $x = (a, b)$ will be called a line and $y = (c, ac + b)$ a point (on that line). (See the remark after Theorem 4.)

What do we gain assuming that $|F_n| = p_n^2$ (for all n)? In this case the field $F = F_n$ has a subfield of $p = p_n$ elements, denoted by G. Let $\alpha \in F$ be a primitive element of F over G. Thus any element in F can be represented in the form $h + s\alpha$ for some $h, s \in G$.

We construct a family of p^3 disjoint p^3-element sets of pairs ⟨a line, a point on that line⟩, whose union covers the set S of all p^6 such pairs. Each set will involve

p^2 different points and p^2 different lines, each of those p^2 points will belong to p of those p^2 lines, and conversely each of those p^2 lines will have p of those p^2 points. To construct such family represent each pair $\langle x, y\rangle \in S$ in the form

$$x = (f + r\alpha, h + s\alpha), \quad y = (g + t\alpha, fg + h + (ft + gr + s)\alpha + rt\alpha^2),$$

where $f, g, h, r, t, s \in G$. Fixing r, t, s we obtain a set S_{rts} of p^3 pairs from S having p^2 lines. Unfortunately S_{rts} has about p^3 points. Let us try to reduce the number of points in each S_{rts}: the substitution $s \mapsto s - ft$ changes the above representation of a pair $\langle x, y\rangle \in S$ to

$$x = (f + r\alpha, h + (s - ft)\alpha), \quad y = (g + t\alpha, fg + h + (gr + s)\alpha + rt\alpha^2).$$

Now, any line in S_{rts} is identified by pair (f, h) thus there are p^2 different lines in S_{rts}; any point in S_{rts} is identified by pair $(g, fg + h)$ thus there are p^2 different points in S_{rts}.

Let us finish the proof. We take as z the set from our family which contains the pair $\langle x, y\rangle$. As the number of sets is p^3 we have $K(z) \leqslant 3 \log p + O(\log n) = 1.5n + O(\log n)$. As each set has p^2 different lines and p^2 different points, we have $K(x|z), K(y|z) \leqslant 2 \log p + O(\log n) = n + O(\log n)$. □

References

[1] P. Gács, J. Körner, Common information is far less than mutual information, Problems Control Inform. Theory 2 (1973) 149–162.

[2] D. Hammer, A. Romashchenko, A. Shen, N. Vereshchagin, Inequalities for Shannon entropies and Kolmogorov complexities, in: Proc. 12th Annu IEEE Conf. on Computational Complexity, Ulm, Germany, June 1997, pp. 13–23 (Final version: Journal of Computer and System Sciences, 60 (2000), p. 442–464).

[3] M. Li, P. Vitányi, An Introduction to Kolmogorov Complexity and its Applications, 2nd ed., Springer, Berlin, 1997.

[4] An.A. Muchnik, On the extraction of common information of two strings, Abstracts of talks at the First World Congress of Bernoulli Society, Moscow, Nauka, 1986, p. 453 (in Russian).

[5] An.A. Muchnik, On common information, Theoret. Comput. Sci 207 (1998) 319–328.

[6] A.E. Romashchenko, Pary slov s nematerialisuemoi vzaimnoi informatsiei (Pairs of strings with no extractable mutual information, in Russian), Problemy peredachi informatsii (Problems Inform. Transmission) 36 (2000) 3–20.

[7] J.R. Shoenfield, Degrees of Unsolvability, North-Holland Publishing Company, Amsterdam, 1971.