NORMALIZED INFORMATION DISTANCE AND THE OSCILLATION HIERARCHY

KLAUS AMBOS-SPIES, WOLFGANG MERKLE, AND SEBASTIAAN A. TERWIJN

Abstract. We study the complexity of computing the normalized information distance. We introduce a hierarchy of limit-computable functions by considering the number of oscillations. This is a function version of the difference hierarchy for sets. We show that the normalized information distance is not in any level of this hierarchy, strengthening previous nonapproximability results. As an ingredient to the proof, we demonstrate a conditional undecidability result about the independence of pairs of random strings.

1. Introduction

Normalized information distance. The normalized information distance NID is a distance measure for binary strings that is based on prefix-free Kolmogorov complexity K. Here the value K(x) is the minimum length of a string p that describes x in the sense that $U(p) = x$ for some fixed additively optimal Turing machine with prefix-free domain. Observe that such a machine cannot be defined on the empty string, hence all values of K are nonzero. The *normalized information distance* is defined as

$$\mathrm{NID}(x,y) = \frac{E(x,y)}{\max\{\mathrm{K}(x),\mathrm{K}(y)\}} \qquad \text{where} \qquad E(x,y) = \max\{\mathrm{K}(x\,|\,y),\mathrm{K}(y\,|\,x)\}.$$

Note that NID, being the ratio of two nonzero functions that are approximable from above, is computable in the limit, i.e., there is a computable rational-valued function f with three arguments such that for all x and y we have

$$\lim_{s\to\infty} f(x,y,s) = \mathrm{NID}(x,y).$$

Terwijn, Torenvliet, and Vitányi [9] have shown that NID can neither be computably approximated from below nor from above, i.e., such a computable approximation f of NID can neither be increasing nor decreasing in s. In particular, the function NID is not computable. In what follows, we improve on these nonapproximability results by confirming their conjecture [9, Section 5] that for any computable approximation of NID, the number of oscillations is not bounded by a constant, or, equivalently, that NID is not in the oscillation hierarchy. The oscillation hierarchy is defined as the union of the classes $\Sigma_1^{-1}, \Sigma_2^{-1}, \ldots$, where Σ_k^{-1} is the class of all functions that have a computable approximation that initially increases and switches at most $k-1$ times between increasing and decreasing. See Section 2 for formal definitions.

Related to the proof of our main result, we demonstrate that given two random strings, it is undecidable whether they are independent. In fact, this conditional undecidability result is derived in the stronger form that there is no enumeration of pairs that includes infinitely many random pairs and where all the random pairs in the enumeration are independent. The stronger result can be viewed as a conditional immunity statement and is used in the proof of our main result.

Date: September 24, 2021.

2010 *Mathematics Subject Classification.* 03D15, 03D32, 03D55, 68Q30.

Key words and phrases. Kolmogorov complexity, information distance, independence.

Related work. The concept of normalized information distance was introduced by Li et al. [4], and subsequently studied in a series of papers, cf. Vitányi et al. [10] and Li and Vitányi [5, Section 8.4]. It has both theoretical and practical interest. While the function NID itself is noncomputable, there are computable variants that have a number of surprising practical applications. Such variants are for example defined in terms of standard compression algorithms in place of prefix-free Kolmogorov complexity.

The difference hierarchy over the computably enumerable sets, or c.e. sets, for short, was introduced by Ershov, cf. Odifreddi [6, IV.1.18] and Selivanov [7]. It is a fine hierarchy for the Δ^0_2-sets, sometimes also referred to as the Boolean hierarchy. It can be seen as an effective version of a classical hierarchy introduced by Hausdorff, which is studied in descriptive set theory. An analogous hierarchy defined over NP is studied in complexity theory. When restricting attention to $\{0,1\}$-valued functions, i.e., to *sets*, the oscillation hierarchy coincides with the difference hierarchy, as follows from the discussion following Definition 2.1. In particular, Σ^{-1}_1 consists of the c.e. sets, Σ^{-1}_2 contains the d.c.e. sets, i.e., the differences of c.e. sets, and in general Σ^{-1}_k contains the k-c.e. sets. These coincidences motivate the choice of our notation for the classes of the oscillation hierarchy, as the same notation has been used for the classes of the difference hierarchy, see e.g. Selivanov [8].

Recall that a hierarchy is proper if each of its levels is strictly included in the next one. Similar to the case of sets, the oscillation hierarchy is proper and does not exhaust the class of all limit-computable functions. As in the case of sets, this can be shown by elementary diagonalization arguments and, in fact, this *follows* from the analogous results for sets. Theorem 5.1, our main result, asserts that NID is a natural example of a limit-computable function that is not in the oscillation hierarchy.

Note that Bennett et al. [2] have shown that E satisfies the properties of a metric up to a constant additive term. Furthermore, E is minimal among all similar distance functions [10, Theorem 3.7]. Note further that somewhat in contrast to the definition of normalized information distance, the name information distance is used for the function D defined as

$$D(x,y) = \min\big\{l(p)\colon U(p,x) = y \wedge U(p,y) = x\big\}.$$

Here U is the universal prefix-free machine used to define K. It can be shown that D and E are equal up to a logarithmic additive term [10, Corollary 3.1], i.e., we have

$$D(x,y) = E(x,y) + O(\log E(x,y)).$$

Notation. Our notation is mostly standard. For further explanations, details and background, in particular about computability theory, we refer to Odifreddi [6] and to Downey and Hirschfeldt [3]. A *string* is a binary word, i.e., a finite sequence over the binary alphabet $\{0,1\}$. We use $|x|$ to denote the *length* of a string x, the empty string λ is the unique string of length 0. The set of strings is denoted by $\{0,1\}^*$, the set of natural numbers is denoted by ω. The two latter sets are identified by the order isomorphism that takes the length-lexicographical ordering on the set of strings to the standard ordering on the natural numbers. Unless explicitly stated differently, the term set refers to a set of binary strings or, equivalently, to a subset of the natural numbers. We identify such a set A with the infinite binary sequence $A(0)A(1)\ldots$ where $A(n)$ is equal to 1 if and only if n is in A. This sequence is referred to as the characteristic sequence of A.

We use $\leqslant^+$ to denote inequality up to a fixed additive constant. For example, $f(x) \leqslant^+ g(x)$ means that there is a constant c such that we have $f(x) \leqslant g(x) + c$

for all x in some specific set that will be clear from the context. Similar notation such as $=^+$ is defined likewise.

Enumerations of any type of objects are always meant to be effective.

Prefix-free Kolmogorov complexity. For further use, we compile some standard facts about Kolmogorov complexity. For proofs of these facts, as well as for definitions, details and further background, we refer to Li and Vitányi [5] and to Downey and Hirschfeldt [3]. For a string x, we let x^* denote the program of minimum length for x that appears first in some fixed enumeration of the domain of the universal machine used to define K. By the latter condition, the string x^* can be computed given x and $\mathrm{K}(x)$. For a string x of length n, we have

$$\mathrm{K}(x) \leqslant^+ n + 2\log n, \tag{1}$$

$$\mathrm{K}(x \,|\, |x|) \leqslant^+ n. \tag{2}$$

Indeed, it holds that $\mathrm{K}(x) \leqslant^+ n + \mathrm{K}(n)$ for all strings x. By Chaitin's counting theorem [3, Theorem 3.7.6], there is a constant d such that for all t for at most a fraction of 2^{-t+d} of all words x of length n, we have $\mathrm{K}(x) \leqslant n + \mathrm{K}(n) - t$. In the special case where t is equal to $\mathrm{K}(n) + 1$, we obtain that at most a fraction of $2^{-\mathrm{K}(n)-1+d}$ of all words x of length n is nonrandom in the sense that $\mathrm{K}(x) < n$. By symmetry of information, we refer to the following chain of equations

$$\begin{aligned} \mathrm{K}(x,y) &=^+ \mathrm{K}(x) + \mathrm{K}(y \,|\, x^*) \\ &=^+ \mathrm{K}(y) + \mathrm{K}(x \,|\, y^*), \end{aligned}$$

which holds for all strings x and y. In case both strings have the same length, we have $\mathrm{K}(xy) =^+ \mathrm{K}(x,y)$, and symmetry of information remains valid with $\mathrm{K}(x,y)$ replaced by $\mathrm{K}(xy)$. Symmetry of information is due to Levin and Gács and also Chaitin [3, Theorem 3.10.2].

Outline. First, in Section 2 we review the limit-computable functions and introduce the oscillation hierarchy. Then, in Section 3, we derive some basic properties of NID and, in particular, reprove the known results that NID can neither be effectively approximated from below nor from above. Before demonstrating in Section 5 our main result, Theorem 5.1, we collect in Section 4 notation and facts to be used in its proof, including the already mentioned conditional immunity result, which is stated as Theorem 4.7.

2. Effective approximations and the oscillation hierarchy

Limit-computable functions. In this section we introduce notation that relates to approximations of real-valued functions on the natural numbers. This notation extends canonically to real-valued functions with a countable domain like $\{0,1\}^*$, $\mathbb{Q}$, $\omega \times \omega$, or similar via the usual identification of such a domain with the set of natural numbers. In particular, this notation will be applied to approximations of the function NID, which maps pairs of strings to a rational number.

For a start, we recall the following notation from computability theory.

Definition 2.1. Let $F\colon \omega \to \mathbb{R}$ be a function. A function $f\colon \omega \times \omega \to \mathbb{Q}$ is an *approximation* of F, if we have for all natural numbers x that

$$\lim_{s\to\infty} f(x,s) = F(x).$$

The function F is *limit computable*, if F has a computable approximation.

Given a computable approximation of an ω-valued function F, by rounding the values of the approximation to the nearest natural number, we obtain a computable ω-valued approximation f to F where then, in particular, for each argument x

almost all values $f(x,s)$ are equal to $F(x)$. As a consequence, ω-valued limit-computable functions are just the limit-computable functions from computability theory, which are also called computably approximable or Δ^0_2-functions. By Shoenfield's Limit Lemma [6, IV.1.17], an ω-valued function F is limit-computable if and only if F is computable with the halting problem $\emptyset'$. The three following remarks show that this equivalence is false for rational-valued functions in general, but it extends to rational-valued functions such as NID, where for given arguments one can compute a finite set of candidate rational numbers that contains the function value. The reason is that for functions of the latter type, from any effective approximation that converges in distance, we obtain an effective approximation that converges in value by rounding the approximating values to the nearest value in the set of candidates, see Remark 2.3 for details in the case of NID.

Remark 2.2. *Let $W_0, W_1, \ldots$ be the standard enumeration of all c.e. sets. Fix some enumeration $(e_0,n_0),(e_1,n_1),\ldots$ of all pairs (e,n) such that n is in W_e, and let $W_{e,s}$ be equal to the set of all n_i such that $e_i = e$ and $i < s$. If we let $F(e) = 1$ in case W_e is empty, let $F(e) = 0$ in case W_e is infinite and, otherwise, let*

$$F(e) = 2^{-\max W_e}, \text{ then } F(e) = \lim_{s\to\infty} f(e,s) \text{ where } f(e,s) = 2^{-\max(W_{e,s}\cup\{0\})}.$$

Note that $F(e)$ is equal to 0 if and only if W_e is infinite. The function f is computable, hence F is limit-computable. However, as a rational function, F is not computable with the halting problem because otherwise, the halting problem would decide the Π^0_2-complete index set of all e such that W_e is infinite, a contradiction.

Remark 2.3. *Let f be a computable approximation of* NID*, i.e., f converges to* NID *in distance in the sense that for any arguments x and y, the difference between $f(x,y,s)$ and $\mathrm{NID}(x,y)$ goes to zero. By definition of* NID *and the upper bounds* (1) *and* (2) *on prefix-free Kolmogorov complexity, for some constant c any value of the form $\mathrm{NID}(x,y)$ must be contained in the set*

$$V(x,y) = \left\{\tfrac{i}{j} \colon i,j \le 2(|x|+|y|)+c\right\}$$

By rounding any value $f(x,y,s)$ to the nearest value in $V(x,y)$, breaking ties arbitrarily, we obtain an approximation f^{R} that converges to NID *not just in distance but also in value, i.e., for all x and y the approximated value $f^{\mathrm{R}}(x,y,s)$ is equal to $\mathrm{NID}(x,y)$ for almost all s.*

Remark 2.4. *A rational-valued function F is computable with the halting problem if and only if it has a computable approximation f that converges to F by value in the sense of Remark 2.3, i.e., such that for all x the value $f(x,s)$ is equal to $F(x)$ for almost all s. The proof is essentially the same as the proof of Shoenfield's Limit Lemma, details are omitted.*

Increasing and decreasing phases. Given an approximation to some limit-computable function, for any given argument we consider the alternations of the approximation between going up and going down. By bounding the number of such changes by a constant for all arguments, we obtain a fine hierarchy for limit-computable functions.

Definition 2.5. Let f be an approximation of some function $\omega \to \mathbb{R}$ and fix some natural number x. With x understood, let

$$\delta(x,s) = f(x,s+1) - f(x,s)$$

be the *increase of f at s*, and call s *increasing* in case $\delta(x,s) > 0$, and call s *decreasing* in case $\delta(x,s) < 0$. Furthermore, a subset of the natural numbers is *monotonic* in case it does not contain both increasing and decreasing indices.

For any given natural number x, *phase t of f on x* is defined inductively for all $t > 0$ as follows. Phase 1 is equal to the maximum initial segment of ω on which f is monotonic. In the induction step, assume that for some $t > 1$ the phases 1 through $t-1$ are already defined. If the union of the latter phases is all of ω, these are the only phases of f on x. Otherwise, let m_t be the maximum member in phase $t-1$ and let phase t be equal to the maximum initial segment of $\omega \setminus \{0, \ldots, m_t\}$ on which f is monotonic.

The approximation f *reaches at most phase t on x* in case there is no phase $t+1$ on x. In case the latter holds for all natural numbers x, the approximation f *reaches at most phase t*. A phase is *increasing* if it contains an increasing index, and a phase is *decreasing* if it contains a decreasing index

The next remark states without proofs some straightforward properties of phases.

Remark 2.6. *Let f be an approximation of some function $\omega \to \mathbb{R}$ and let x be a natural number. The phases of f on x form a partition of the natural numbers into successive contiguous intervals, which are all finite unless the partition is finite, in which case exactly the last phase is infinite. In case the function $s \mapsto f(x,s)$ is constant, there is exactly one phase, which is neither increasing nor decreasing. Otherwise, each phase is either increasing or decreasing, and increasing and decreasing phases alternate. With the possible exception of phase 1, a phase is increasing or decreasing if and only if the least index in the phase is increasing or decreasing, respectively.*

The oscillation hierarchy. The levels of the oscillation hierarchy introduced next stratify the class of limit-computable functions according to the number of alternations between increasing and decreasing phases.

Definition 2.7. Let k be a nonzero natural number. A Σ_k^{-1}-approximation is a computable approximation f of some function $\mathbb{N} \to \mathbb{R}$ such that on every input the approximation in the first phase is either constant or increasing and f reaches at most phase k. The definition of Π_k^{-1}-approximation is literally the same except that the first phase is required to be decreasing instead of increasing.

A function $F\colon \omega \to \mathbb{Q}$ is a Σ_k^{-1}-*function* in case it has a Σ_k^{-1}-approximation. The class of all Σ_k^{-1}-functions is denoted by Σ_k^{-1}. The notion of a Π_k^{-1}-*function* and the class Π_k^{-1} of all such functions is defined likewise. The *oscillation hierarchy* is defined as

$$\bigcup_{k\geq 1} \left(\Sigma_k^{-1} \cup \Pi_k^{-1}\right) = \bigcup_{k\geq 1} \Sigma_k^{-1} = \bigcup_{k\geq 1} \Pi_k^{-1}.$$

The functions in Σ_1^{-1} and in Π_1^{-1} are also called *approximable from below* and *approximable from above*, respectively.

Normalizing approximations of NID. We write $\mathrm{NID}_s(x,y)$ for approximations of NID, i.e., we have

$$\lim_{s\to\infty} \mathrm{NID}_s(x,y) = \mathrm{NID}(x,y) \quad \text{and} \quad \mathrm{NID}_s(x,y) \in \mathbb{Q}.$$

Notions relating to approximations are extended to this notation in the natural way, e.g., such an approximation is computable if $\mathrm{NID}_s(x,y)$ is a computable function of s, x, and y. In the same fashion, let K_s be some fixed computable approximation from above to K_s with values in the natural numbers, and similar for conditional prefix-free Kolmogorov complexity.

Definition 2.8. The *Kolmogorov approximation* $\mathrm{NID}_s^{\mathrm{K}}$ to NID is defined by

$$\mathrm{NID}_s^{\mathrm{K}}(x,y) = \frac{\max\left\{\mathrm{K}_s(x\,|\,y), \mathrm{K}_s(y\,|\,x)\right\}}{\max\left\{\mathrm{K}_s(x), \mathrm{K}_s(y)\right\}}.$$

Remark 2.9. *Let* NID_s *be any effective approximation of* NID *that reaches at most phase* m *and, like in Remark 2.3, let* $\mathrm{NID}_s^{\mathrm{R}}$ *be the version of* NID_s *where the function values have been rounded to the nearest value in the set* $V(x,y)$*. Then for all* x *and* y *and for almost all* i*, we have*

$$\mathrm{NID}_i^{\mathrm{R}}(x,y) = \mathrm{NID}_i^{\mathrm{K}}(x,y) \tag{3}$$

because both sides of the equation converge in value to $\mathrm{NID}(x,y)$ *in the sense of Remark 2.3. So we can fix a computable sequence* $i(0) < i(1) < \dots$ *such that* (3) *holds with* i *replaced by* $i(s)$ *for all* s*. Let*

$$\mathrm{NID}'_s(x,y) = \mathrm{NID}^{\mathrm{R}}_{i(s)}(x,y) = \mathrm{NID}^{\mathrm{K}}_{i(s)}(x,y)$$

and call NID'_s *the* normalized version *of* NID_s*. Note that* NID'_s *is indeed an effective approximation of* NID *and reaches at most phase* m*, too. For a proof of the latter property, observe that* $\mathrm{NID}_i^{\mathrm{R}}$ *reaches at most phase* m *since the latter function may only increase in* s *in case* NID_s *increases, and similarly for decreasing. Thus it suffices to observe that by construction* $\mathrm{NID}'_0(x,y), \mathrm{NID}'_1(x,y), \dots$ *is a subsequence of* $\mathrm{NID}_0^{\mathrm{R}}(x,y), \mathrm{NID}_1^{\mathrm{R}}(x,y), \dots$*.*

3. Some basic properties of NID

In Section 5, we will show our main result that NID is not in the oscillation hierarchy. Before, we derive in the current section some basic properties of NID and give new proofs for the known facts [9] that NID is approximable from neither below nor above.

Lemma 3.1. *The values of* $\mathrm{NID}(x,y)$ *come arbitrarily close to* 0 *and* 1 *even if the arguments are restricted to strings* x *and* y *of the same length. In fact, the following slightly stronger assertions hold*

$$\lim_{n\to\infty} \max\{\mathrm{NID}(x,x) \,:\, |x| = n\} = 0, \tag{4}$$

$$\lim_{n\to\infty} \max\{\mathrm{NID}(x,0^n) \colon |x| = n\} = 1. \tag{5}$$

Proof. For a proof of (4), observe that by definition we have $\mathrm{NID}(x,x) = \frac{\mathrm{K}(x\,|\,x)}{\mathrm{K}(x)}$. For fractions of the latter form, with growing length of x the denominator goes to infinity, whereas the numerator is bounded from above by a constant, so $\mathrm{NID}(x,x)$ tends to 0. For a proof of (5), observe that by a standard counting argument, for some constant c, all sufficiently large n and some x of length n, we have

$$\mathrm{K}(0^n\,|\,x) \le n - c \le \mathrm{K}(x\,|\,0^n) \qquad \text{and} \qquad \mathrm{K}(0^n) \le \mathrm{K}(x) \le n + 3\log n.$$

So by definition of NID and (1), it holds for almost all n and all such x that

$$\mathrm{NID}(x,0^n) = \frac{\mathrm{K}(x\,|\,0^n)}{\mathrm{K}(x)} \ge \frac{n-c}{n+3\log n} \;\longrightarrow_{n\to\infty}\; 1.$$

□

A set is called *immune* if it is infinite, but it does not contain an infinite c.e. subset. Immune sets were introduced by Post, and they play an important role in computability theory, cf. Odifreddi [6].

Theorem 3.2. (Bārzdiņš) *The set* $\{x \colon \mathrm{K}(x) \geqslant \frac{1}{2}|x|\}$ *is immune.*

Proof. In case the theorem were false, fix an enumeration of some infinite c.e. subset of the set under consideration. Among all strings of length at least t, let x_t be the one that is enumerated first. There is a prefix machine with some coding constant c that outputs x_{4n} when given the string 10^{n-1} as input, hence $2n \le \mathrm{K}(x_{4n}) \le n + c$ for all n, a contradiction. □

Proposition 3.3. *Let r be a real number where $0 < r < 1$. Then the set*

$$X_r = \{(x, y) \colon |x| = |y| \text{ and } \mathrm{NID}(x, y) > r\}$$

is immune.

Proof. First note that X is infinite by Lemma 3.1. Now suppose for a contradiction that X_r has an infinite c.e. subset A. For each n there are at most finitely many pairs (x, y) where $|x| = |y| = n$, hence by taking an appropriate effective subsequence of some fixed enumeration of A, we obtain an enumeration $(x_0, y_0), (x_1, y_1), \ldots$ of some infinite c.e. subset of A where $|x_n| < |x_{n+1}|$ for all n. By the latter property and because x_n and y_n have equal length by definition of X_r, the values $\max\{\mathrm{K}(x_n), \mathrm{K}(y_n)\}$ tend to infinity, while the values $\mathrm{K}(x_n \,|\, y_n)$ and $\mathrm{K}(y_n \,|\, x_n)$ are both bounded from above by a fixed constant that does not depend on n. Consequently, the values $\mathrm{NID}(x_n, y_n)$ tend to 0, a contradiction. □

Theorem 3.4. ([9]) NID *is not approximable from below.*

Proof. By Proposition 3.3, the set $X_{1/3}$ defined there is immune. But if NID were approximable from below, this set would be c.e., hence could not be immune. □

Lemma 3.5. *There is no computable sequence of pairs (x_k, y_k) such that $|x_k| = |y_k|$ and $\mathrm{NID}(x_k, y_k) < \frac{1}{k}$ for all k.*

Proof. Assume for a contradiction that there is a sequence as in the lemma. Fix a constant c_0 such that for all strings x and y of equal length, the values $\mathrm{K}(x)$ and $\mathrm{K}(y)$ are both less than or equal to $\mathrm{K}(xy) + c_0$. There is a prefix-free machine with some coding constant c_1 that outputs $x_{2k}y_{2k}$ when given the binary string 10^{k-1} as input, hence $\mathrm{K}(x_{2k}y_{2k}) \leq k + c_1$ for all k. In summary, we have

$$\frac{1}{2k} > \mathrm{NID}(x_{2k}, y_{2k}) \geq \frac{1}{\max\{\mathrm{K}(x_{2k}), \mathrm{K}(y_{2k})\}} \geq \frac{1}{\mathrm{K}(x_{2k}y_{2k}) + c_0} \geq \frac{1}{k + c_0 + c_1},$$

a plain contradiction for all $k > c_0 + c_1$. □

Proposition 3.6. ([9]) NID *is not approximable from above.*

Proof. Assume for a proof by contradiction that the proposition is false. By Lemma 3.1, the values $\mathrm{NID}(x, x)$ tend to 0, thus by dovetailing approximations from above to the values $\mathrm{NID}(x, x)$ for all x, for given k one can effectively find a string x_k such that $\mathrm{NID}(x_k, x_k) < \frac{1}{k}$. This contradicts Lemma 3.5. □

Theorem 3.7. NID *is not a Σ_2^{-1}-function.*

Proof. Suppose for a contradiction that NID is Σ_2^{-1}, that is, it has a computable approximation that starts with an increasing phase and reaches at most phase 2. Consider the pairs (x, y) of words where $|x| = |y|$. By Lemma 3.1, there are infinitely many such pairs (x, y) where $\mathrm{NID}(x, y) > \frac{3}{4}$. Consequently, we can effectively find infinitely many such pairs (x, y) such that the approximation of $\mathrm{NID}(x, y)$ attains a value strictly larger than $\frac{3}{4}$ during phase 1. If for some $k \geq 2$ and almost all pairs (x, y) of the latter kind it would actually hold that $\mathrm{NID}(x, y) > \frac{1}{k}$ this would contradict Proposition 3.3. As a consequence, for every $k \geq 2$ there is a pair (x, y) of words of identical length where the approximation becomes smaller than $\frac{1}{k}$ during phase 2, and for all such k, x, and y we have $\mathrm{NID}(x, y) < \frac{1}{k}$ because the approximation never reaches phase 3. For given k such x and y can be found effectively, which contradicts Lemma 3.5. □

4. Conditional independence

Random and independent pairs. Before we demonstrate in the next section our main result, Theorem 5.1, we collect some notation and facts used in its proof.

Definition 4.1. Let $r > 0$ be a real number and let a and a' be words. The word a is *random* if $\mathrm{K}(a) \geq |a|$, and the pair (a, a') is *random* if a and a' are both random.

The string a is *r-compressible* if $\mathrm{K}(a) \leq r|a|$, and the pair (a, a') is *r-compressible* if a and a' are both r-compressible. The pair (a, a') is *mutually r-compressible* if we have

$$\mathrm{K}(a \,|\, a') \leq r|a| \quad \text{and} \quad \mathrm{K}(a' \,|\, a) \leq r|a'|. \tag{6}$$

Lemma 4.2. *Let $\varepsilon > 0$ be a real number. For almost all n, all but a fraction of at most ε of the pairs (a, a') of words of equal length n are random.*

Proof. By Chaitin's counting theorem, there is a constant d such that for given n, at most $2^{n-\mathrm{K}(n)-1+d} \cdot 2^n$ many pairs have a nonrandom first component, and the same bound holds for the number of pairs with nonrandom second component. Consequently, among the 2^{2n} pairs of words of length n at most $2^{2n-\mathrm{K}(n)+d}$ are nonrandom, which is a fraction of at most ε for almost all n. □

As usual, let an *order* be a function from the set of natural numbers to itself that is nondecreasing and unbounded.

Definition 4.3. With some computable order h understood, a pair (a, a') of strings of equal length n is *independent conditioned on a string x* if we have

$$\mathrm{K}(a \,|\, a'^*, x^*) \geq |a| - h(n) \quad \text{and} \quad \mathrm{K}(a' \,|\, a^*, x^*) \geq |a'| - h(n), \tag{7}$$

and (a, a') is *independent* if it is independent conditioned on the empty string.

Lemma 4.4. *Let $\varepsilon > 0$ be a real number and let h be some order. Then there is some n_0 such that for all $n \geq n_0$ and for any fixed word x, all but a fraction of at most ε of the pairs (a, a') of words of equal length n are independent conditioned on x.*

Proof. Fix any natural number n and any word x. The number of words of length strictly less than $n - h(n)$ is bounded from above by $2^{n-h(n)}$, hence for given a' the latter bounds also the number of words a such that $\mathrm{K}(a \,|\, a'^*, x^*) < n - h(n)$. As a consequence, the number of pairs (a, a') of words of length n that do not satisfy the first inequality in (7) is at most $2^n 2^{n-h(n)}$. By symmetry, the same upper bound holds for the number of pairs (a, a') that do not satisfy the second inequality in (7). Consequently, among the 2^{2n} pairs of words of length n, at most $2 \cdot 2^{2n-h(n)}$ many pairs are not independent conditioned on x, i.e., at most a fraction of $2 \cdot 2^{-h(n)}$. The latter bound is at most ε for all n larger than some appropriate number n_0 that does not depend on x. □

Conditional immunity. As a further ingredient to the proof of Theorem 5.1, we derive a result about the undecidability of independence of random strings. More precisely, we show that there is no algorithm that, given two random strings of the same length, can decide whether they are independent or not, where it is agreed that the algorithm may fail to converge or to give the right answer if one or both of the strings are not random. In fact, we need a stronger assertion, which will be formulated in terms of the following notion of conditional immunity.

Definition 4.5. A set A is *decidable conditional to* a set C if there is a partial computable function φ such that for all x in C the value $\varphi(x)$ is defined and equal to $A(x)$. A set A is *immune conditional to* a set C if the set $A \cap C$ is infinite but does not contain an infinite set of the form $B \cap C$ where the set B is c.e.

Suppose that A is decidable conditional to C and that this is witnessed by the partial computable function φ. Then A cannot be immune conditional to C because the set A is either finite or contains the infinite set $B \cap C$ where B is the c.e. set of all n where $\varphi(n)$ is equal to 1. Hence, conditional immunity is a strong form of conditional undecidability.

Decidability and immunity conditional to the set of natural numbers are just classical decidability and immunity, respectively. Classically, a set is not immune if it is either finite or has an infinite c.e. subset. Indeed, this c.e. subset can always be assumed to decidable, since every infinite c.e. set contains a decidable subset. The following remark shows that a similar remark is false for conditional immunity.

Remark 4.6. *There are sets A and C with infinite intersection such that A is not immune conditional to C but the latter fact is not witnessed by any computable set.*

For a proof, let D be any infinite subset of some c.e. set A, and let $C = D \cup \overline{A}$. Then A witnesses that A is not immune conditional to C. Moreover, for any set B that witnesses the latter fact, by definition the set $B \cap C$ is an infinite subset of A, hence B is a subset of A that has an infinite intersection with D.

So it suffices to show that any noncomputable c.e. set A has an infinite subset D where $B \cap D$ is finite for any computable subset B of A. To this end, let $B_0, B_1, \ldots$ be a (noneffective) list of all computable subsets of A and let $F_e = \bigcup_{i \leqslant e} B_i$ for all e. Each set F_e is a computable subset of A, hence $A \setminus F_e$ is infinite by noncomputability of A. This implies that there is an ascending sequence $d_0 < d_1 < \ldots$ of elements of A such that $d_e \notin F_e$ for all e. So the set $D = \{d_e \colon e \geqslant 0\}$ is an infinite subset of A such that for all e, the intersection $D \cap B_e$ is contained in $\{d_0, \ldots, d_{e-1}\}$.

Theorem 4.7. *Let $r > 0$ be a real number. Let R be the set of random pairs of equal length and let I be the set of pairs of equal length that are not mutually r-compressible, i.e., let*

$$R = \big\{(x,y) \colon |x| = |y| \wedge \mathrm{K}(x) \geqslant |x| \wedge \mathrm{K}(y) \geqslant |y|\big\},$$
$$I = \big\{(x,y) \colon |x| = |y| \wedge \big(\mathrm{K}(x \,|\, y) > r|x| \vee \mathrm{K}(y \,|\, x) > r|y|\big)\big\}.$$

Then the set I is immune conditional to R.

Proof. By the discussion of Chaitin's counting theorem in the introduction and Lemma 4.2, it follows easily that the intersection of the sets R and I is infinite, details are left to the reader. So suppose for a contradiction that there exists a c.e. set B such that $R \cap B$ is an infinite subset of I. Fix any pair (x,y) in $R \cap B$ and without loss of generality assume $\mathrm{K}(y \,|\, x) \geqslant r|y|$. If we let n be equal to the length of x and y, we have for some constant c

$$\begin{aligned}
\mathrm{K}(xy) &\geqslant^{+} \mathrm{K}(x) + \mathrm{K}(y \,|\, x^*) \\
&\geqslant^{+} \mathrm{K}(x) + \mathrm{K}(y \,|\, x, \mathrm{K}(x)) \\
&\geqslant^{+} n + \mathrm{K}(y \,|\, x) - c \log n \\
&\geqslant \quad n + rn - c \log n.
\end{aligned}$$

Here the inequalities follow, from top to bottom, by the variant of symmetry of information stated in the paragraph on Kolmogorov complexity, because x^* can be computed given x and $\mathrm{K}(x)$, because applying (1) twice yields $\mathrm{K}(\mathrm{K}(x)) < c \log n$ for some constant c, and, finally, by assumption on the pair (x,y).

Consider any n that is so large that $c \log n < \frac{r}{2} n$, and such that $R \cap B$ contains pairs (x,y) of words of length n. Then, on the one hand, for each such pair, we have $\mathrm{K}(xy) \geq^{+} n + \frac{r}{2} n$. On the other hand, for each such n there is such a pair (x_n, y_n) where $\mathrm{K}(x_n y_n) \leqslant^{+} n$, a contradiction. In order to obtain (x_n, y_n) as claimed, let z_n be the string of length n that is enumerated last in some fixed enumeration of all nonrandom strings (of all lengths). Then knowing z_n one knows

all random strings of length n. Thus we can compute from z_n the pair (x_n, y_n) that among all random pairs of strings of length n is enumerated first into B. Since $\mathrm{K}(z_n) < n$, we have $\mathrm{K}(x_n y_n) \leqslant^+ n$. □

5. NID is not in the oscillation hierarchy

Our main result Theorem 5.1 asserts that NID is not in the oscillation hierarchy, which confirms a conjecture by Terwijn, Torenvliet, and Vitányi [9].

We begin by giving an informal description of the proof of Theorem 5.1. For a proof by contradiction, we assume that there is a computable approximation NID_s to NID that reaches at most phase m for some natural number m. By Remark 2.9, we can assume that this approximation NID_s is normalized, i.e., is obtained by approximating prefix-free Kolmogorov complexity. We may thus argue, for example, that the approximated values $\mathrm{NID}_s(x, y)$ become larger in case the approximations to $\mathrm{K}(x)$ and $\mathrm{K}(y)$ become smaller while the approximations to $\mathrm{K}(y \mid x)$ and $\mathrm{K}(y \mid x)$ remain the same. By using such formulations we aim at a very rough intuitive description of the phenomena that occur, which is, however, not precise enough to provide a sketch of the formal proof.

In the proof of Theorem 5.1, we fix rational numbers α and β where $\beta < \alpha < 1$. The proof has an inductive structure where in the induction step we consider approximations $\mathrm{NID}_s(w, w')$ for pairs of strings $w = abc$ and $w' = a'b'c'$ where a and a', b and b', as well as c and c' are of identical length, and where a has length n, b has length $2n$, and c has length ℓn for some fixed ℓ where $6 \leq \ell \leq 3^m - 3$. We

use an independence condition for pairs of words where the fraction of pairs that do not satisfy the condition among all pairs of words of length ℓn tends to zero when n goes to infinity. Thus if some property holds for almost all n and a constant nonzero fraction of all pairs (c, c') of words of length ℓn, then for almost all n and some slightly smaller constant nonzero fraction of all such pairs, both the property and the independence condition hold.

In the induction step, we assume that there is an increasing phase t_0 during which the approximation goes above α for infinitely many pairs (a, a') and some constant nonzero fraction of all pairs $(bc, b'c')$. Then we argue that this includes infinitely many pairs $(ab, a'b')$ such that at some later stage the pair (b, b') appears to be at the same time random and mutually highly compressible. By the latter property and the independence condition it follows that $\mathrm{NID}(abc, a'b'c') < \beta$, which in turn implies that there must be a decreasing phase $t_1 > t_0$ during which the approximation goes below β for infinitely many pairs $(ab, a'b')$ and some constant nonzero fraction of all pairs (c, c'). Next we argue that for infinitely many of these pairs $(ab, a'b')$ it turns out later that the pair (b, b') is mutually highly compressible, which together with the independence condition implies that $\mathrm{NID}(abc, a'b'c') > \alpha$. Consequently, there must be an increasing phase $t_2 > t_1$ during which the approximation goes above α for infinitely many pairs $(ab, a'b')$ and a nonzero fraction of all pairs (c, c').

Intuitively speaking, in the induction step it is argued that there are sufficiently many argument pairs $(abc, a'b'c')$ for which the approximation NID_s first goes above α during phase t_0, then goes below β during phase t_1, and finally goes again above α during phase t_2. This holds because there are sufficiently many pairs b and b' that first appear to be random and mutually incompressible, then, second, appear to be random and mutually compressible, and, third, finally appear to be nonrandom and mutually compressible. That is, the maximum of $\mathrm{K}(b)$ and

of $\mathrm{K}(b')$ and the maximum of $\mathrm{K}(b \,|\, b')$ and $\mathrm{K}(b' \,|\, b)$ appear first to be both high, second to be high and low, respectively, and, third, to be both low, where low means close to 0 and high means close to $|b|$. That such changes, which concern only the strings b and b', result in changes of the value of $\mathrm{NID}_s(abc, a'b'c')$ depends on the notion of independence. For an independent pair (c, c'), the prefix-free Kolmogorov complexity of c and c', as well as their mutual conditional prefix-free Kolmogorov complexity conditioned in addition on $(ab)^*$ are all so close to $|c|$ that the influence of c on a value of the form $\mathrm{NID}_s(abc, a'b'c')$ can be neglected compared to the influence of a, a', b, and b'. Since in addition the two former strings are short compared to the two latter strings, the described changes in prefix-free Kolmogorov complexity relating to b and b', though small compared to $|c|$, are still large enough to force $\mathrm{NID}_s(abc, a'b'c')$ below β and above α.

Theorem 5.1. NID *is not in the oscillation hierarchy, i.e.,* NID *is not in* Σ_m^{-1} *for any* $m \geqslant 1$.

Proof. For a proof by contradiction, assume that NID is in Σ_m^{-1} for some $m > 1$, i.e., has a computable approximation $\mathrm{NID}_s(x, y)$ that always starts with an increasing phase and reaches at most phase m on all arguments. In what follows, we speak of increasing phases $1, 3, \ldots$ and decreasing phases $2, 4, \ldots$. Choose the rational $r > 0$ so small that

$$\alpha := \frac{1 - 5r}{1} \quad \text{is strictly larger than} \quad \beta := \frac{3^m - 2 + 4r}{3^m - 1}.$$

For the scope of this proof, call a pair (w, w') of words t-high in case phase t is increasing and contains some s where $\mathrm{NID}_{s+1}(w, w') > \alpha$. Similarly, call the pair t-low in case phase t is decreasing and contains some s where $\mathrm{NID}_{s+1}(w, w') < \beta$. Given natural numbers k and t, and a real number ε, let

$$A(k, t, \varepsilon) = \{(a,a') \colon |a| = |a'| \text{ and for a fraction of at least } \varepsilon \text{ of all pairs } (u, u') \\ \text{of words of equal length } (3^k - 1)|a|, \text{ the pair } (au, a'u') \text{ is } t\text{-high}\}.$$

Observe that all sets of the form $A(k, t, \varepsilon)$ are empty in case $t > m$, as well as in case phase t is decreasing, by choice of NID_s and by definition of t-high.

In the remainder of this proof, the notion independent conditioned on a certain word is always meant with respect to the fixed order $h(n) = \log n$. In particular, the values $h(n)/n$ tend to 0, hence for any constant ℓ, we have

$$\frac{h(\ell n)}{n} = \ell \, \frac{h(\ell n)}{\ell n} \xrightarrow{n \to \infty} 0.$$

Claim 1. *There is some phase* $t \leq m$ *such that* $A(m, t, \frac{1}{2m})$ *is infinite.*

Proof. Let n be a natural number, let $a = 0^n$ and let u and u' be any words of length $(3^m - 1)n$. Then we have

$$\begin{aligned} \mathrm{K}(au) &=^+ \mathrm{K}(u), \\ \mathrm{K}(au') &=^+ \mathrm{K}(u'), \\ \mathrm{K}(au \,|\, au') &=^+ \mathrm{K}(u \,|\, u'), \\ \mathrm{K}(au' \,|\, au) &=^+ \mathrm{K}(u' \,|\, u), \end{aligned}$$

where the constants hidden in the notation $=^+$ do not depend on n, a, u or u'. Thus for some constant d that is again independent of the latter four parameters,

in case n is sufficiently large and u and u' are independent, we have

$$\begin{aligned}\mathrm{NID}(au, au') &\geq \frac{\max\{\mathrm{K}(u \,|\, u'), \mathrm{K}(u' \,|\, u)\} - d}{\max\{\mathrm{K}(u), \mathrm{K}(u')\} + d} \\ &\geq \frac{|u| - h(|u|) - d}{|u| + \mathrm{K}(|u|) + 2d} \\ &\geq \frac{|u| - 2\log|u|}{|u| + 3\log|u|} \\ &\geq \frac{|u| - 5\log|u|}{|u|} > \alpha.\end{aligned}$$

By the preceding discussion and Lemma 4.4, for almost all n and at least half of all pairs (u, u') of words of length $(3^m - 1)n$, we have $\mathrm{NID}(0^n u, 0^n u') > \alpha$, hence $(0^n u, 0^n u')$ must be t'-high for some phase t', where $t' \leq m$ by assumption on the approximation NID_s. Hence there must be some $t \leq m$ such that for infinitely many n for a fraction of at least $\frac{1}{2m}$ of all pairs (u, u') of words of length $(3^m - 1)n$ the pair $(0^n u, 0^n u')$ is t-high. For all such n, the pair $(0^n, 0^n)$ is in $A(m, t, \frac{1}{2m})$. □

Claim 2. *Let k and t be in $\{2, \ldots, m\}$, and let $\varepsilon > 0$ be a real number such that $A(k, t, \varepsilon)$ is infinite. Then $A(k-1, t', \frac{\varepsilon}{4m^2})$ is infinite for some $t' \geq t + 2$.*

Before we prove Claim 2, we argue that the first two claims imply the theorem. By Claim 1, we can fix $t \leq m$ such that $A(m, t, \frac{1}{2m})$ is infinite. By applying Claim 2 to the latter set for at most $\lceil \frac{m}{2} \rceil$ times, we obtain $j \in \{1, \ldots, \lceil \frac{m}{2} \rceil\}$, $\widetilde{t} > m$, and $\widetilde{\varepsilon} > 0$ such that the set $A(m - j, \widetilde{t}, \widetilde{\varepsilon})$ is infinite. This is a contradiction because the latter set must be empty as the approximation NID_s is assumed to reach at most phase $m < \widetilde{t}$. Observe that $m - j \geq m - \lceil \frac{m}{2} \rceil \geq 1$ since $m > 1$.

In order to demonstrate Claim 2, fix k, t_0 and ε as in the assumption of the claim. For the remainder of this proof, when using the letters a, b, c, w, and n with or without decoration in the same context, we always assume that we have

$$w = abc, \quad |a| = n, \quad |b| = 2n, \quad |c| = \ell n \quad \text{where } \ell = 3^k - 3, \text{ i.e., } |w| = 3^k n.$$

In particular, we assume for all pairs of the form (a, a'), $(ab, a'b')$, or similar that the two components of the pair have equal length. By abuse of notation, quantification over words and pairs of words involving the mentioned variable names is restricted to words of the form just described. For example, if we use the phrase for all words a and b, this is meant as abbreviating the phrase for all n and all words a of length n and b of length $2n$. Recall the concept of a random pair introduced in Definition 4.1.

Claim 3. *Let the pair (c, c') be independent conditioned on $(ab, a'b')$ where the pair (b, b') is random and mutually r-compressible. Then we have*

$$\mathrm{NID}(abc, a'b'c') < \beta.$$

Proof. The assumption of the claim implies that

$$\mathrm{K}(abc) =^+ \mathrm{K}(ab) + \mathrm{K}(c \,|\, (ab)^*) \geq^+ |b| + |c| - h(|c|) \geq^+ (\ell + 2)n - h(\ell n),$$

where the equation holds by symmetry of information, and the first inequality holds because b is random and (c, c') is independent conditioned on $(ab, a'b')$. By symmetry, the derived lower bound also holds for $\mathrm{K}(a'b'c')$. Furthermore, we have

$$\begin{aligned}\mathrm{K}(abc \,|\, a'b'c') &\leq^+ \mathrm{K}(ab \,|\, a'b'c') + \mathrm{K}(c \,|\, a'b'c') \leq^+ \mathrm{K}(ab \,|\, a'b') + \mathrm{K}(c \,|\, n) \\ &\leq^+ |a| + r|b| + |c| = (1 + 2r + \ell)n,\end{aligned}$$

where by symmetry again, this upper bound also holds for $\mathrm{K}(a'b'c' \,|\, abc)$. By the lower and upper bounds just derived, there is a constant d such that for all sufficiently large n we have

$$\begin{aligned}\mathrm{NID}(abc, a'b'c') &= \frac{\max\{\mathrm{K}(abc \,|\, a'b'c'), \mathrm{K}(a'b'c' \,|\, abc)\}}{\max\{\mathrm{K}(abc), \mathrm{K}(a'b'c')\}} \leq \frac{(\ell+1+2r)n+d}{(\ell+2)n - h(\ell n) - d} \\ &= \frac{\ell+1+2r+d/n}{\ell+2-h(\ell n)/n - d/n} < \frac{\ell+1+3r}{\ell+2-r} < \frac{3^k-2+4r}{3^k-1} \leq \beta.\end{aligned}$$

□

Claim 4. *Let (c, c') be independent conditioned on $(ab, a'b')$ and let the pair $(ab, a'b')$ be r-compressible. Then $\mathrm{NID}(abc, a'b'c') > \alpha$.*

Proof. For all sufficiently large n, we have

$$\begin{aligned}\mathrm{K}(abc) &\leq^+ \mathrm{K}(ab) + \mathrm{K}(c \,|\, (ab)^*) \leq^+ r|ab| + \mathrm{K}(c \,|\, n) \leq^+ (3r+\ell)n \\ \mathrm{K}(abc \,|\, a'b'c') &\geq^+ \mathrm{K}(c \,|\, (a'b')^* c'^*) \geq^+ \ell n - h(\ell n),\end{aligned}$$

where by symmetry the upper bound holds also for $\mathrm{K}(a'b'c')$ and the lower bound holds also for $\mathrm{K}(a'b'c' \,|\, abc)$. Similar to the proof of Claim 3, we obtain that there is a constant d such that for all sufficiently large n we have

$$\mathrm{NID}(abc, a'b'c') \geq \frac{\ell n - h(\ell n) - d}{(\ell+3r)n + d} = \frac{\ell - h(\ell n)/n - d/n}{\ell + 3r + d/n} > \frac{\ell - r}{\ell + 4r} > \frac{\ell - 5r}{\ell} > \alpha$$

□

Claim 5. *Infinitely many pairs $(ab, a'b')$ where the pair (b, b') is random and mutually r-compressible are member of the set*

$$B_0 = \{(ab, a'b') \colon (abc, a'b'c') \text{ is } t_0\text{-high for a fraction of at least } \varepsilon/2 \text{ of all } (c, c')\}.$$

Proof. For any pair (a, a') in $A(k, t_0, \varepsilon)$, the pair $(ab, a'b')$ is in B_0 for a fraction of at least $\varepsilon/2$ of all pairs (b, b'). Otherwise, if this fraction were $q < \varepsilon/2$, the fraction of pairs $(bc, b'c')$ such that $(abc, a'b'c')$ is t_0-high would be strictly less than $q + (1-q)\frac{\varepsilon}{2} < \varepsilon$, contrary to the definition of $A(k, t_0, \varepsilon)$. By Lemma 4.2, for almost all n all but a fraction of $\varepsilon/4$ of all pairs (b, b') are random. So for almost all of the infinitely many (a, a') in $A(k, t_0, \varepsilon)$, there is some $(ab, a'b')$ in B_0 where the pair (b, b') is random.

For given s, w and w', one can compute the value of $\mathrm{NID}_s(w, w')$ and the phase in which s is, hence the set B_0 is c.e. But then the set of all pairs (b, b') such that $(ab, a'b')$ is in B_0 for some words a and a' is also c.e. By the discussion in the last paragraph, the latter c.e. set contains infinitely many random pairs, and then infinitely many of these random pairs must be mutually r-compressible by Theorem 4.7. □

Claim 6. *There is a decreasing phase $t_1 > t_0$ such that the set B_1 is infinite, where*

$$B_1 = \{(ab, a'b') \colon (abc, a'b'c') \text{ is } t_1\text{-low for a fraction of at least } \tfrac{\varepsilon}{3m} \text{ of all } (c, c')\}.$$

Proof. By Lemma 4.4, for almost all n and for any given words a, a', b, and b', all but a fraction of $\varepsilon/6$ of the pairs (c, c') are independent conditioned on $(ab, a'b')$, hence almost all pairs $(ab, a'b')$ in B_0 are a member of the set

$$\begin{aligned}B_0' = \{(ab, a'b') \colon &(abc, a'b'c') \text{ is } t_0\text{-high and } (c, c') \text{ is independent conditioned} \\ &\text{on } (ab, a'b') \text{ for a fraction of at least } \varepsilon/3 \text{ of all pairs } (c, c')\}.\end{aligned}$$

The definition of the set B_0' is meant such that the conditions on being t_0-high and being independent must be satisfied simultaneously for the specified fraction of all pairs (c, c'), and this convention is extended to subsequent similar formulations.

Since the set B_0 is a subset of B_0' except for finitely many members of B_0, Claim 5 holds with B_0 replaced by B_0'. Thus by Claim 3, the set

$$B_0'' = \{(ab, a'b') \colon (abc, a'b'c') \text{ is } t_0\text{-high and } \mathrm{NID}(abc, a'b'c') < \beta \text{ for a fraction of at least } \varepsilon/3 \text{ of all pairs } (c, c')\}$$

is infinite, too. The phase t_0 is increasing, hence for a pair $(abc, a'b'c')$ that is t_0-high but has NID-value of less than β, i.e., where we have

$$\mathrm{NID}(abc, a'b'c') < \beta < \alpha < \mathrm{NID}_{t_0}(abc, a'b'c'),$$

there must be some decreasing phase $t > t_0$ such that the pair is t-low. Since by assumption there are at most m phases, the claim follows. □

Claim 7. *There is an increasing phase $t_2 > t_1$ such that the set B_2 is infinite where*

$$B_2 = \{(ab, a'b') \colon (abc, a'b'c') \textit{ is } t_2\textit{-high for at least } \tfrac{\varepsilon}{4m^2} \textit{ of all (c, c')}\}.$$

Proof. The proof is very similar to the proof of Claim 6 and we omit details that are obvious by this similarity. The set B_1 is infinite and c.e. Then B_1 must contain infinitely many pairs that are r-compressible because otherwise for almost all pairs $(ab, a'b')$ in B_1 we would have $\mathrm{K}(aba'b') \geq \frac{r}{3}|aba'b'|$. The latter contradicts a straightforward variant of Theorem 3.2, which follows by essentially the same proof as the theorem.

Furthermore, by essentially the same argument as in the case of B_0 and B_0' it follows that almost all pairs in B_1 are also in the set

$$B_1' = \{(ab, a'b') \colon (abc, a'b'c') \text{ is } t_1\text{-low and } (c, c') \text{ is independent conditioned on } (ab, a'b') \text{ for a fraction of at least } \tfrac{\varepsilon}{4m} \text{ of all pairs } (c, c')\}.$$

By the preceding discussion, the set B_1' contains infinitely many pairs $(ab, a'b')$ where $(abc, a'b'c')$ is t_1-low and the assumption of Claim 4 is satisfied for a fraction of at least $\varepsilon/4m$ of all pairs (c, c'), hence the set

$$B_1'' = \{(ab, a'b') \colon (abc, a'b'c') \text{ is } t_1\text{-low and } \mathrm{NID}(abc, a'b'c') > \alpha \text{ for a fraction of at least } \tfrac{\varepsilon}{4m} \text{ of all pairs } (c, c')\}$$

is infinite, too. The phase t_1 is decreasing, hence for a pair $(abc, a'b'c')$ that is t_1-low but has NID-value greater than α, there must be some increasing phase $t_2 > t_1$ such that the pair is t_2-high. By assumption there are at most m phases, the claim follows. □

Now Claim 2 follows because we have $t_2 > t_1 > t_0 > 0$ and the set B_2 is equal to $A(k-1, t_2, \frac{\varepsilon}{4m^2})$ since we have $|ab| = 3n$ and $|c| = (3^k - 3)n = (3^{k-1} - 1)3n$. □

From the proof of Theorem 5.1 it is obvious that the examples of pairs of strings x,y forcing the changes in the approximation NID_s are of rather long length. It would be interesting to have a more careful analysis of these lengths.[1]

Question 5.2. *Relate the number of oscillations of approximations of* $\mathrm{NID}(x, y)$ *to the length of x and y.*

Acknowlegements

We are grateful to an anonymous referee for helpful comment and corrections, which include a remark addressing Question 5.2: by the proof of the main result the number of required oscillations must be at least logarithmic in the length of the considered strings.

[1]Question 5.2 is addressed in the paper by Bauwens and Blinnikov [1], which appeared after the arxiv version of the current paper from August 11, 2017.

References

[1] B. F. Bauwens and I. Blinnikov, *The normalized algorithmic information distance can not be approximated*, in: Computer Science – Theory and Applications, 15th International Computer Science Symposium in Russia, CSR 2020, Yekaterinburg, Russia, Proceedings Vol. 12159, Springer (2020) 130–141.

[2] C. H. Bennett, P. Gács, M. Li, P. M. B. Vitányi, and W. Zurek, *Information distance*, IEEE Transactions on Information Theory 44 (1998) 1407–1423.

[3] R. G. Downey and D. R. Hirschfeldt, *Algorithmic Randomness and Complexity*, Springer-Verlag, 2010.

[4] M. Li, X. Chen, X. Li, B. Ma, P. Vitányi, *The similarity metric*, IEEE Transactions on Information Theory, 50(12) (2004) 3250–3264.

[5] M. Li and P. M. B. Vitányi, *An introduction to Kolmogorov Complexity and Its Applications*, third edition, Springer-Verlag, 2008.

[6] P. Odifreddi, *Classical Recursion Theory*, Vol. 1, Studies in Logic and the Foundations of Mathematics Vol. 125, North-Holland, 1989.

[7] V. L. Selivanov, *Fine hierarchies and Boolean terms*, Journal of Symbolic Logic 60 (1995) 289–317.

[8] V. L. Selivanov, *Difference hierarchy in ϕ-spaces*, Algebra and Logic 43(4) (2004) 238–248.

[9] S. A. Terwijn, L. Torenvliet, and P. M. B. Vitányi, *Nonapproximability of the normalized information distance*, Journal of Computer and System Sciences 77 (2011) 738–742.

[10] P. M. B. Vitányi, F. J. Balbach, R. Cilibrasi, M. Li, *Normalized information distance*, pp. 45-82 in: Information Theory and Statistical Learning, F. Emmert-Streib and M. Dehmer (eds.), Springer-Verlag, 2008.

(Klaus Ambos-Spies) Universität Heidelberg, Institut für Informatik, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

Email address: ambos@math.uni-heidelberg.de

(Wolfgang Merkle) Universität Heidelberg, Institut für Informatik, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany

Email address: merkle@math.uni-heidelberg.de

(Sebastiaan A. Terwijn) Radboud University Nijmegen, Department of Mathematics, P.O. Box 9010, 6500 GL Nijmegen, the Netherlands.

Email address: terwijn@math.ru.nl