Математические структуры и моделирование
201X, вып. 2x, с. 1–??

УДК 000.000

NEGATIVE RESULTS OF COMPUTABLE ANALYSIS DISAPPEAR IF WE RESTRICT OURSELVES TO RANDOM (OR, MORE GENERALLY, TYPICAL) INPUTS

V. Kreinovich

It is well known that many computational problems are, in general, not algorithmically solvable: e.g., it is not possible to algorithmically decide whether two computable real numbers are equal, and it is not possible to compute the roots of a computable function. We propose to constraint such operations to certain "sets of typical elements" or "sets of random elements".

In our previous papers, we proposed (and analyzed) physics-motivated definitions for these notions. In short, a set $\mathcal{T}$ is a *set of typical elements* if for every definable sequences of sets A_n with $A_n \supseteq A_{n+1}$ and $\bigcap\limits_n A_n = \emptyset$, there exists an N for which $A_N \cap \mathcal{T} = \emptyset$; the definition of a *set of random elements* with respect to a probability measure P is similar, with the condition $\bigcap\limits_n A_n = \emptyset$ replaced by a more general condition $\lim\limits_n P(A_n) = 0$.

In this paper, we show that if we restrict computations to such typical or random elements, then problems which are non-computable in the general case – like comparing real numbers or finding the roots of a computable function – become computable.

1. Negative results of computable analysis: a brief reminder

Physically meaningful computations with real numbers: a brief reminder. In practice, many quantities such as weight, speed, etc., are characterized by real numbers. To get information about the corresponding value x, we perform measurements. Measurements are never absolute accurate. As a result of each measurement, we get a measurement result $\widetilde{x}$; for each measurement, we usually also know the upper bound Δ on the (absolute value of) the measurement error $\Delta x \stackrel{\text{def}}{=} \widetilde{x} - x$: $|x - \widetilde{x}| \le \Delta$.

To fully characterize a value x, we must measure it with a higher and higher accuracy. As a result, when we perform measurements with accuracy 2^{-n} with $n = 0, 1, \ldots$, we get a sequence of rational numbers r_n for which $|x - r_n| \le 2^{-n}$.

Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968, USA
E-mail: vladik@utep.edu

From the computational viewpoint, we can view this sequence as an "oracle" (subroutine) that, given an integer n, returns a rational number r_n. Such sequences represent real numbers in computable analysis; see, e.g., [13, 15].

Remark 1. When the mapping from n to r_n is algorithmic, then the real number is called *computable*.

Meaning of computability: reminder. We say that an algorithm produces a real number z if this algorithm, given n, produces a rational number r_n which is 2^{-n}-close to z. Similarly, we say that an algorithm takes a real number x as an input if this algorithm can, after computing some auxiliary value m, get a rational number which is 2^{-m}-close to x – and use it in further computations.

For example, there exists an algorithm that, given two real numbers x and y, computes their sum $z = x + y$. Indeed, to compute the desired 2^{-n}-approximation z_n to $z = x + y$, it is sufficient to compute $2^{-(n+1)}$-approximations x_{n+1} to x and y_{n+1} to y and add them up.

In this case, we use $m = n + 1$. To compute a product $z = x \cdot y$, we need to use more complex values m.

First negative result. In computable analysis, several negative results are known. For example, it is known that no algorithm is possible that, given two real numbers x and y, would check whether these numbers are equal or not.

Functions: from computability viewpoint. Similarly, we can define a (uniformly) continuous function $f(x)$ from real numbers to real numbers as a mapping that, given:

- an integer n (describing the desired accuracy of $f(x)$),
- an integer m (describing the accuracy with which we know the input x, and
- a 2^{-m}-accurate rational approximation r to the input x,

produces either a rational number y_n which is guaranteed to be 2^{-n}-close to $f(x)$, or a message that the given input accuracy m is insufficient for such computations.

We say that an algorithm uses f as an input if this algorithm, after generating the auxiliary values n, m, and r, can get either y_n or the message, and use this output in its computations. For example, if the result is a message, then it makes sense to find a more accurate approximation to x (e.g., a $2^{-(m+1)}$-approximation) and try again.

We can similarly define functions $f(x_1, \ldots, x_k)$ of several real variables – and what it means for an algorithm that uses such functions as an input.

Remark 2. When this mapping is computable, we get a notion of a *computable function*.

Metric spaces: computational viewpoint. Real numbers are not the only possible physical objects, we can have fields, operators, etc. What is a computational meaning of such a more general object?

On each stage, we only have a finite amount of information about each object. If this object is a number, we only know a few first digits in its expansion. If this is a field, we know its values at different points, maybe the average values over some region – in total, finite number of bits. When the number of bits is limited by an integer B, we can have no more than $1+2+\ldots+2^B < 2^{B+1}$ possible combinations of such bits – and thus, finitely many possible approximations to the actual object. By combining the approximating objects corresponding to all possible values B, we get a potentially infinite sequence of approximations $x_1, \ldots, x_n, \ldots$ such that each object can be, with any given accuracy, be approximated by such objects. To describe the accuracy of such an approximation, we need to know, for every two approximations x_i and x_j, the distance $d(x_i, x_j)$ between them.

Thus, by a metric space, in computable analysis, we usually understand a mapping that, given three integers i, j, and n, returns a rational 2^{-n}-approximation d_n to the distance $d(x_i, x_j)$. We say that an algorithm takes a metric space as an input if this algorithm, after producing auxiliary values of i, j, and n, can get such a rational approximation d_n.

From the mathematical viewpoint, this means that we consider *separable* metric spaces, i.e., metric spaces in which there is a sequence $\{x_n\}$ such that every element from the metric space can be approximated, with any given accuracy, by an element of this sequence.

From this viewpoint, to describe an element $x \in X$ of a metric space, we must be able, given n, to produce a 2^{-n}-approximation to x. In other words, having an element x means having a mapping that, given an integer n, produces an integer m for which $d(x, x_m) \le 2^{-n}$. We say that an algorithm takes elements of a metric space as inputs if this algorithm, after generating an auxiliary number n, can get such m. This is a generalization of what it means to have a real number as an input.

From the physical viewpoint, it is sufficient to consider complete metric spaces. To explain this point, let us first go back to computable real numbers. Let us assume that someone has proposed a physical theory in which only rational values of time (as measured in some fixed time unit) are possible. Is there a way to experimentally check this theory? For example, is it possible to experimentally check that the moment $\sqrt{2}$ is impossible?

No, because in practice, as we have mentioned, we only observe objects with some accuracy. Within a fixed accuracy, every number can be approximated by a rational one, and thus, no matter how many measurements we make, we will never be able to tell whether the actual value is rational or irrational – because both values are consistent with all the measurements. Because of this, without changing consistency with observations, we can always safely assume that all real numbers are physically possible.

In general, we can similarly safely assume that all limits of the points x_i are possible, i.e., that we have a sequence x_{n_k} that converges, i.e., for which $d(x_{n_k}, d_{n_l}) \to$

0 as $k, l \to \infty$, then it has a limit in the metric space X. In mathematical terms, such metric spaces are called *complete*. So this means that from the physical viewpoint, it is sufficient to only consider metric spaces.

Compact spaces from the viewpoint of computable analysis. It is known that a closed set K in a complete metric space X is compact if and only if it has a finite ε-net for every real number $\varepsilon > 0$, i.e., it has a finite set $\{y_1, \ldots, y_m\}$ such that for every $x \in K$, there exists an element y_j for which $d(x, y_j) \le \varepsilon$. One can easily check that it is sufficient to require this property for values $\varepsilon = 2^{-n}$, $n = 0, 1, \ldots$ One can also easily check that in separable metric spaces, we can always replace the the elements y_j by their approximations from the sequence x_i.

It is therefore reasonable, from the computational viewpoint, to describe a compact as a mapping that, given an integer n, returns a finite list of natural numbers $i_1, \ldots, i_m$ such that for every $x \in K$, there exists a $j \le m$ for which $d(x, y_{i_j}) \le 2^{-n}$.

Remark 3. If this mapping – as well as the mapping that produce $d(x_i, x_j)$ – are algorithmic, then we have a *computable compact*.

Negative results about computations with functions. Several negative results are known about computations with functions. For example,

- while there is an algorithm that, given a function $f(x)$ on a compact set K (e.g., on a box $[\underline{x}_1, \overline{x}_1] \times \ldots \times [\underline{x}_k, \overline{x}_k]$ in k-dimensional space), produces the values $\max\{f(x) : x \in K\}$,
- no algorithm is possible that would always return a point x at which this maximum is attained (and similarly, with minimum).

2. Negative results of computable analysis: the physicists' viewpoint

From the physicists' viewpoint, these negative results seem rather theoretical. From the purely mathematical viewpoint, if two quantities coincide up to 13 digits, they may still turn to be different: for example, they may be 1 and $1 + 10^{-100}$.

However, in the physics practice, if two quantities coincide up to a very high accuracy, it is a good indication that they are actually equal. This is how physical theories are confirmed: if an experimentally observed value of a quantity turned out to be very close to the value predicted based on a theory, this means that this theory is (triumphantly) true. This is, for example, how General Relativity has been confirmed.

This is how discoveries are often made: for example, when it turned out the speed of the waves described by Maxwell equations of electrodynamics is very close to the observed speed of light c, this led physicists to realize that light is formed of electromagnetic waves.

How physicists argue. A typical physicist argument is that while numbers like $1 + 10^{-100}$ (or $c \cdot (1 + 10^{-100})$) are, in principle, possible, they are abnormal (not typical).

When a physicist argues that second order terms like $a \cdot \Delta x^2$ of the Taylor expansion can be ignored in some approximate computations because Δx is small, the argument is that

- while abnormally high values of a (e.g., $a = 10^{40}$) are mathematically possible,
- typical (= not abnormal) values appearing in physical equations are usually of reasonable size.

How to formalize the physicist's intuition of typical (not abnormal). A formalization of this intuition was proposed and analyzed in [3, 4, 7–11]. Its main idea is as follows. To some physicist, all the values of a coefficient a above 10 are abnormal. To another one, who is more cautious, all the values above 10 000 are abnormal. Yet another physicist may have another threshold above which everything is abnormal. However, for every physicist, there is a value n such that all value above n are abnormal.

This argument can be generalized as a following property of the set $\mathcal{T}$ of all typical elements. Suppose that we have a monotonically decreasing sequence of sets $A_1 \supseteq A_2 \supseteq \ldots$ for which $\bigcap\limits_n A_n = \emptyset$ (in the above example, A_n is the set of all numbers $\geq n$). Then, there exists an integer N for which $\mathcal{T} \cap A_N = \emptyset$.

We thus arrive at the following definition:

Definition 1. We say that $\mathcal{T}$ is a *set of typical elements* if for every definable decreasing sequence $\{A_n\}$ for which $\bigcap\limits_n A_n = \emptyset$, there exists an N for which

$$\mathcal{T} \cap A_N = \emptyset.$$

Remark 4. The word "definable" is understood in the usual way. Let $\mathcal{L}$ be a theory, let $P(x)$ be a formula from the language of the theory $\mathcal{L}$, with one free variable x so that the set $\{x \,|\, P(x)\}$ is defined in $\mathcal{L}$. We will then call the set $\{x \,|\, P(x)\}$ *$\mathcal{L}$-definable*.

Our objective is to be able to make mathematical statements about $\mathcal{L}$-definable sets. Thus, we must restrict definability to a *subset* of properties, so that the resulting notion of definability will be defined in whatever language we use. In other words, we must have a stronger theory $\mathcal{M}$ in which the class of all $\mathcal{L}$-definable sets is a countable set. One can prove that such $\mathcal{M}$ always exists; for details, see, e.g., [7].

In the following proofs, we will assume that $\mathcal{L}$ is ZFC, and so, any definition which works in ZFC leads to a definable object.

For example, a metric space is usually defined as a pair of a set X and a metric function $d : X \times X \to \mathbb{R}$ with usual properties. If both are definable – i.e., if both are uniquely determined by some closed formulas of ZFC – we call the corresponding metric space definable.

The notion of definability may sound somewhat evasive. Every example that we can come up with, be it a set or a real number, or a metric space, is definable. This does not mean, of course, that every real number is definable: there are countably many formulas and thus, countably many definable numbers, but there are more than countably many real numbers.

One can prove that the above definition of the set of typical elements is consistent in the following sense:

Theorem 1. *For every probability measure P on the universal set X which is defined on all definable subsets of X, and for every real number $\varepsilon > 0$, there exists a set $\mathcal{T}$ of typical elements for which the lower probability is $\geq 1 - \varepsilon$: $\underline{P}(\mathcal{T}) \geq 1 - \varepsilon$.*

Proof. There are countably many monotonically decreasing definable sequences with empty intersection: $\{A_n\}$: $\{A_n^{(1)}\}$, $\{A_n^{(2)}\}$, ... For each k, since the sequence is monotonically decreasing and has an empty intersection, we have $P\left(A_n^{(k)}\right) \to 0$ as $n \to \infty$. Hence, there exists N_k for which $P\left(A_{N_k}^{(k)}\right) \leq \varepsilon \cdot 2^{-k}$. We can now take $\mathcal{T} \stackrel{\text{def}}{=} - \bigcup\limits_{k=1}^{\infty} A_{N_k}^{(k)}$. Since $P\left(A_{N_k}^{(k)}\right) \leq \varepsilon \cdot 2^{-k}$, we have

$$\overline{P}\left(\bigcup_{k=1}^{\infty} A_{N_k}^{(k)}\right) \leq \sum_{k=1}^{\infty} P\left(A_{N_k}^{(k)}\right) \leq \sum_{k=1}^{\infty} \varepsilon \cdot 2^{-k} = \varepsilon.$$

Hence, $\underline{P}(\mathcal{T}) = 1 - \overline{P}\left(\bigcup\limits_{k=1}^{\infty} A_{N_k}^{(k)}\right) \geq 1 - \varepsilon$. ■

Relation to randomness. The above notion of typicality is related to the notion of a random object (see, e.g., [12]).

Namely, Kolmogorov and Martin-Löf proposed a new definition of a random sequence, a definition that separates physically random binary sequences – e.g., sequences that appear in coin flipping experiments or sequences that appear in quantum measurements – from sequence that follow some pattern. Intuitively, if a sequence s is random, it satisfies all the probability laws – like the law of large numbers, the central limit theorem, etc. Vice versa, if a sequence satisfies all probability laws, then for all practical purposes we can consider it random. Thus, we can define a sequence to be random if it satisfies all probability laws.

What is a probability law? In precise terms, it is a statement S which is true with probability 1: $P(S) = 1$. So, to prove that a sequence is not random, we must show that it does not satisfy one of these laws.

Equivalently, this statement can be reformulated as follows: a sequence s is *not* random if $s \in C$ for a (definable) set C $(= -S)$ with $P(C) = 0$. As a result, we arrive at the following definition:

Definition 2. We say that a sequence is *random* if it does not belong to any definable set of measure 0.

(If we use different languages to formalize the notion "definable", we get different versions of Kolmogorov-Martin-Löf randomness.)

It is easy to prove that this definition is consistent, in the sense that almost all sequences are random. Indeed, every definable set C is defined by a finite sequence of symbols (its definition). Since there are countably many sequences of symbols, there are (at most) countably many definable sets C. So, the complement $-\mathcal{R}$ to the class $\mathcal{R}$ of all random sequences also has probability 0.

Informally, this definition means that (definable) events with probability 0 cannot happen. In practice, physicists also assume that events with a *very small* probability cannot happen. For example, they believe that it is not possible that all the molecules in the originally uniform air move to one side of the room – although, from the viewpoint of statistical physics, the probability of this event is not zero. This fits very well with a commonsense understanding of rare events: e.g., if a coin falls head 100 times in a row (or a casino roulette gets to red 100 times in a row), any reasonable person will conclude that this coin is not fair.

It is not possible to formalize this idea by simply setting a threshold $p_0 > 0$ below which events are not possible – since then, for N for which $2^{-N} < p_0$, no sequence of N heads or tails would be possible at all. However, we know that for each monotonic sequence of properties A_n with $\lim p(A_n) = 0$ (e.g., $A_n =$ "we can get first n heads"), there exists an N above which a truly random sequence cannot belong to A_N. In [3, 4, 7–11], we thus propose the following definition of a set of random elements:

Definition 3. We say that $\mathcal{R}$ is a *set of random elements* if for every definable decreasing sequence $\{A_n\}$ for which $\lim P(A_n) = 0$, there exists an N for which

$$\mathcal{R} \cap A_N = \emptyset.$$

Let us show, on the example of coin tossing, that this definition indeed formalizes our intuition. In this case, the universal set is the set of all sequences of Heads (H) and Tails (T): $U = \{\mathrm{H}, \mathrm{T}\}^{\mathbb{N}}$ Here, A_n is the set of all the sequences that start with n heads. The sequence $\{A_n\}$ is decreasing and definable, and its intersection has probability 0. Therefore, for every set $\mathcal{R}$ of random elements of U, there exists an integer N for which $A_N \cap \mathcal{R} = \emptyset$. This means that a random sequence cannot start with N heads. This is exactly what we wanted to formalize.

The above definition is very similar to the definition of the set of typical elements; the only difference is that the condition $\bigcap\limits_n A_n = \emptyset$ is replaced with a more general condition $\lim P(A_n) = 0$. This relation leads to the following relation between these two definitions. Let $\mathcal{R}_K$ denote the set of the elements random in the usual Kolmogorov-Martin-Löf sense. Then the following is true [4]:

Theorem 2.

- *Every set of random elements is also a set of typical elements.*
- *For every set of typical elements $\mathcal{T}$, the intersection $\mathcal{T} \cap \mathcal{R}_K$ is a set of random elements.*

Proof. If $\cap A_n = \emptyset$ then $P(A_n) \to 0$. Thus, every set of random elements is also a set of typical elements.

Vice versa, let $\mathcal{T}$ be a set of typical elements. Let us prove that $\mathcal{T} \cap \mathcal{R}_K$ is a set of random elements. Indeed, if $P(\cap A_n) = 0$ then for $B_m \stackrel{\text{def}}{=} A_m - \cap A_n$, we have $B_m \supseteq B_{m+1}$ and $\cap B_n = \emptyset$. Thus, by definition of a set of typical elements, we conclude that there exists an integer N for which $B_N \cap \mathcal{T} = \emptyset$). Since $P(\cap A_n) = 0$, we also know that $(\cap A_n) \cap \mathcal{R}_K = \emptyset$. Thus, $A_N = B_N \cup (\cap A_n)$ has no common elements with the intersection $\mathcal{T} \cap \mathcal{R}_K$. ■

Restriction to typical elements does not mean restriction to a finite grid. Of course, it is possible to take a finite set as $\mathcal{T}$ – this final set satisfies all the properties of a set of typical elements. However, we can also have these sets as large as possible. For example, as we have mentioned, for every $\varepsilon > 0$, there is a set of random elements on the interval $[0, 1]$ which has measure $\geq 1 - \varepsilon$ – and this set is also a set of typical elements.

Physically interesting consequences of these definitions. These definitions have useful consequences [3, 4, 7–11].

Ill-posed problems. The first example is related to *inverse problems* (see, e.g., [14]). These problems are related to the main objectives of science: to provide (ideally) *guaranteed* estimates for physical quantities and (ideally) *guaranteed* predictions for these quantities. The problem with getting such guarantees is that estimation and prediction are *ill-posed problems* in the sense that very small changes in the measurement results can lead to very large changes in the reconstructed state.

One reason for this phenomenon is that measurement devices are inertial. As a result, they suppress high frequencies ω in the measured signal. For such high frequencies ω, signals $\varphi(t)$ and $\varphi(t) + A \cdot \sin(\omega \cdot t)$ are indistinguishable. So, based on the measurements only, we cannot tell whether the actual signal is the original signal $\varphi(t)$ or – for some large A – a very difference signal $\varphi(t) + A \cdot \sin(\omega \cdot t)$.

The existing approaches to this problem are based on some prior assumptions about the actual signal. For example, if we know the actual probability distribution on the set of all possible signals, we can get *statistical* regularization (filtering). If we know bounds on the actual signal's rate of change, e.g., if we know that $|\dot{\varphi}| \leq \Delta$ for some known Δ, then we can use *Tikhonov regularization*. Experts can provide other information about the actual signal, in which case we have expert-based regularization. The problem is that we rarely have this information. We may assume some bounds on the rate of change – but then there is no guarantee that the prediction based on this assumption is correct.

In precise terms, this problem can be formulated as follows. Let S denote the set of all possible states, and let R denote the set of all possible measurement results. In this description, an (ideal) measurement is a continuous 1-1 mapping $f : S \to R$. In principle, we can reconstruct the original state s from the measurement result $r = f(s)$ by applying the inverse function $s = f^{-1}(r)$. However, the inverse function

is, in general, not continuous. As a result, very small measurement errors (changes in r) can lead to drastic changes in the reconstructed state $f^{-1}(r)$. It turns out that if we take into account that the actual states should be typical (i.e., belong to a set of typical states), then this problem disappears.

Definition 4. A definable metric space S is called *definably separable* if there exists a definable sequence $s_1, \ldots, s_n, \ldots$ which is everywhere dense in S.

Theorem 3. *Let S be a definably separable metric space, let $\mathcal{T}$ be a set of all typical elements of S, and let $f : S \to R$ be a continuous 1-1 function. Then, the inverse mapping $f^{-1} : R \to S$ is continuous for every $r \in f(\mathcal{T})$.*

Proof. It is well known that if a f is continuous and 1-1 on a compact, then the inverse function f^{-1} is also continuous. So, to prove our result, it is sufficient to prove that the set $\mathcal{T}$ is *pre-compact*, i.e., that its closure is compact. In a metric space, a set X is compact if and only if it is closed and for every ε, it has a finite ε-net.

Since the metric space S is definably separable, there exists a definable sequence $s_1, \ldots, s_n, \ldots$ which is everywhere dense in S. Let us take take $A_n \stackrel{\text{def}}{=} -\bigcup_{i=1}^{n} B_\varepsilon(s_i)$. Since s_i are everywhere dense, we have $\cap A_n = \emptyset$. Hence, there exists N for which $A_N \cap \mathcal{T} = \emptyset$. Since $A_N = -\bigcup_{i=1}^{N} B_\varepsilon(s_i)$, this means $\mathcal{T} \subseteq \bigcup_{i=1}^{N} B_\varepsilon(s_i)$. Hence $\{s_1, \ldots, s_N\}$ is an ε-net for $\mathcal{T}$. ■

This continuity means that, in contrast to general ill-defined problem, if we perform measurements accurately enough, we can reconstruct the state of the system with any desired accuracy.

Justification of physical induction. What is physical induction? This is a conclusion that physicists make: if a property P is satisfied in the first N experiments, and this number N is sufficiently large, then the property is satisfied always. It turns out that our definition enables us to formalize this idea.

For every property P and for every system o, we can define a sequence of values $s \stackrel{\text{def}}{=} s_1 s_2 \ldots$, where:

- $s_i = T$ if P holds in the i-th experiment with the system o, and
- $s_i = F$ if $\neg P$ holds in the i-th experiment with the system o.

Let X be the set of all such sequences. It is reasonable to require that if the system is typical, then the resulting sequence s is also typical i.e., belongs to some set $\mathcal{T}$ of typical elements.

Theorem 4. *For every set of typical element $\mathcal{T} \subseteq X$, there exists an integer N such that if or some $s \in \mathcal{T}$, we have $s_1 = \ldots = S_N = T$, then $s_m = T$ for all m.*

Proof. Let us consider the following sequence of sets:

$$A_n \stackrel{\text{def}}{=} \{o : s_1 = \ldots = s_n = T \,\&\, \exists m\, (s_m = F)\}.$$

Once can easily check that $A_n \supseteq A_{n+1}$ and $\cap A_n = \emptyset$. Thus, there exists an integer N for which $A_N \cap \mathcal{T} = \emptyset$. This means that if $s \in \mathcal{T}$ and $s_1 = \ldots = s_N = T$, then we cannot have m for which $s_m = F$, i.e., $s_m = T$ for all m. ■

In other words, there exists an N such that if for a typical sequence, a property is satisfied in the first N experiments, then it is satisfied always – this is exactly physical induction.

3. New results: when we restrict ourselves to typical elements, algorithms become possible

In this paper, we analyze the computability consequences of the above definitions. Specifically, we show that most negative results of computability analysis disappear if we restrict ourselves to typical elements.

Deciding equality. Our first result is about checking whether two given real numbers are equal or not, the problem which, as we have mentioned, is, in general, algorithmically unsolvable.

Theorem 5. *For every set of typical pairs of real numbers $\mathcal{T} \subseteq \mathbb{R}^2$, there exists an algorithm, that, given real numbers $(x, y) \in \mathcal{T}$, decides whether $x = y$ or not.*

Proof. The main idea behind this proof is that we can take the sets

$$A_n = \{(x, y) : 0 < d(x, y) < 2^{-n}\}.$$

Then, we have $A_n \supseteq A_{n+1}$ and $\cap A_n = \emptyset$, so there exists an integer N for which $A_N \cap \mathcal{T} = \emptyset$, i.e., for which, if $(x, y) \in \mathcal{T}$, then either $d(x, y) = 0$ (i.e., $x = y$) or $d(x, y) \geq 2^{-N}$.

Thus, if we know that the pair (x, y) belongs to the set $\mathcal{T}$, we can decide whether $x = y$ by using the following algorithm. We compute $d(x, y)$ with accuracy $2^{-(N+2)}$, i.e., compute d such that $|d(x, y) - d| \leq 2^{-(N+2)}$. Then:

- if $d \geq 2^{-(N+1)}$, then $d(x, y) \geq d - 2^{-(N+2)} \geq 2^{-(N+1)} - 2^{-(N+2)} > 0$, hence $x \neq y$;
- if $d < 2^{-(N+1)}$, then $d(x, y) \leq d + 2^{-(N+2)} \leq 2^{-(N+1)} + 2^{-(N+2)} < -2^{-N}$, hence $x = y$.

■

Remark 5. This result is trivially true if $\mathcal{T}$ is a finite set. What we have shown is that a deciding algorithm is possible even when the set $\mathcal{T}$ contains "practically all" pairs – e.g., if on a square $[0, 1] \times [0, 1]$, we take the set $\mathcal{T}$ of measure $\geq 1 - \varepsilon$ for a very small $\varepsilon > 0$.

Remark 6. This and following results are similar to results of Matheiu Hoyrup on layerwise computability (see, e.g., [2]: we have computability on each set $\mathcal{T}$, and – as we have mentioned earlier – we can have such sets with probability $P(\mathcal{T}) \geq 1-\varepsilon$, for any given ε.

Finding roots. As we have mentioned, in general, it is not possible, given a computable function, to compute its root. This becomes possible if we restrict ourselves to *typical* functions:

Theorem 6. *Let K be a computable compact, and let X be the set of all functions $f : K \to \mathbb{IR}$ that attains 0 value somewhere on K. Then, for every set of typical elements $\mathcal{T} \subseteq X$, there exists an algorithm that, given a function $f \in \mathcal{T}$, computes a point x at which $f(x) = 0$.*

Moreover, we can not only produce a root x, we can actually compute, for any given n, an 2^{-n}-approximations to the corresponding set of roots $\{x : f(x) = 0\}$ in the sense of the Hausdorff distance

$$d_H(A, B) \stackrel{\text{def}}{=} \max\left(\sup_{a\in A} d(a, B), \sup_{b\in B} d(b, A)\right),$$

where $d(a, B) \stackrel{\text{def}}{=} \inf\limits_{b\in B} d(a, b)$.

Remark 7. In other words, there exists an algorithm that, given a typical function $f(x)$ on a computable compact K that attains a 0 value somewhere on K, computes a root x – and also computes an 2^{-n}-approximations to the corresponding set of roots.

Proof. To compute the set $R \stackrel{\text{def}}{=} \{x : f(x) = 0\}$ with accuracy $\varepsilon > 0$, let us take an $(\varepsilon/2)$-net $\{x_1, \ldots, x_n\} \subseteq K$. Such a net exists, since K is a computable compact set.

For each i, we can compute $\varepsilon' \in (\varepsilon/2, \varepsilon)$ for which $B_i \stackrel{\text{def}}{=} \{x : d(x, x_i) \leq \varepsilon'\}$ is a computable compact set; see, e.g., [1]. It is possible to algorithmically compute the maximum of a function on a computable compact set; thus, we can compute the value $m_i \stackrel{\text{def}}{=} \min\{|f(x)| : x \in B_i\}$. Since $f \in \mathcal{T}$, similarly to the previous proof, we can prove that there exists an integer N for which, for all $f \in \mathcal{T}$ and for all i, we have either $m_i = 0$ or $m_i \geq 2^{-N}$. Thus, by computing each m_i with accuracy $2^{-(N+2)}$, we can check whether $m_i = 0$ or $m_i > 0$.

We claim that $d_H(R, \{x_i : m_i = 0\}) \leq \varepsilon$, i.e., that:

- for every point x_i for which $m_i = 0$, there exists an ε-close root x, and
- for every root x, there exists an ε-close point x_i for which $m_i = 0$.

Indeed, if $m_i = 0$, this means that the minimum of a function $|f(x)|$ on the ε'-ball B_i with a center in x_i is equal to 0. Since the set K is compact, this value 0 is attained, i.e., there exists a value $x \in B_i$ for which $f(x) = 0$. From $x \in B_i$, we

conclude that $d(x, x_i) \le \varepsilon'$ and, since $\varepsilon' < \varepsilon$, that $d(x, x_i) < \varepsilon$. Thus, x_i is ε-close to the root x.

Vice versa, let x be a root, i.e., let $f(x) = 0$. Since the points x_i form an $(\varepsilon/2)$-net, there exists an index i for which $d(x, x_i) \le \varepsilon/2$. Since $\varepsilon/2 < \varepsilon'$, this means that $d(x, x_i) \le \varepsilon'$ and thus, $x \in B_i$. Therefore, $m_i = \min\{|f(x)| : x \in B_i\} = 0$. So, the root x is ε-close to a point x_i for which $m_i = 0$. ■

Computing fixed points. In general, it is not possible, given a function from a compact set to itself, to compute its fixed point. This becomes possible if we restrict ourselves to *typical* functions:

Theorem 7. *Let K be a computable compact, and let X be the set of all functions $f : K \to K$ that have a fixed point x for which $f(x) = x$. Then, for every set of typical elements $\mathcal{T} \subseteq X$, there exists an algorithm that, given a function $f \in \mathcal{T}$, computes a point x at which $f(x) = x$.*

Moreover, we can not only produce such a fixed point, we can actually compute, for any given n, an 2^{-n}-approximation to the corresponding set of all fixed points $\{x : f(x) = x\}$.

Remark 8. In other words, there exists an algorithm that, given a typical function $f(x)$ on a computable compact K that has a fixed point, computes this fixed point – and also computes an 2^{-n}-approximations to the corresponding set of all fixed points.

Proof. This problem can be reduced to the root finding problem if we take into consideration that that $f(x) = x$ if and only if $g(x) = 0$, where $g(x) \stackrel{\text{def}}{=} d(f(x), x)$. ■

Locating global maxima. In general, it is not possible, given a computable function, to find a point where it attains its maximum. This becomes possible if we restrict ourselves to *typical* functions:

Theorem 8. *Let K be a computable compact, and let X be the set of all functions $f : K \to \mathbb{IR}$. Then, for every set of typical elements $\mathcal{T} \subseteq X$, there exists an algorithm that, given a function $f \in \mathcal{T}$, computes a point x at which $f(x) = \max\limits_{y} f(y)$.*

Moreover, we can not only produce such a point, we can actually compute, for any given n, an 2^{-n}-approximations to the corresponding set of global maximum locations $\left\{x : f(x) = \max\limits_{y} f(y)\right\}$.

Remark 9. In other words, there exists an algorithm that, given a typical function $f(x)$ on a computable compact K, computes a point where this function attains its maximum – and also computes an 2^{-n}-approximations to the corresponding set of all such points.

Proof. This problem can be reduced to previous one if we take into consideration the fact that maximum $\max\limits_y f(y)$ of a computable function on a computable compact is computable and that that $f(x) = \max\limits_y f(y)$ if and only if $g(x) = 0$, where $g(x) \stackrel{\text{def}}{=} f(x) - \max\limits_y f(y)$. ■

Computing minimax strategies. Is it similarly possible to compute the optimal *minimax* strategies, i.e., find x such that

$$\min_y f(x, y) = \max_z \min_y f(z, y).$$

Indeed, this problem is equivalent to finding location of the maximum of a computable function $g(x) \stackrel{\text{def}}{=} \min\limits_y f(x, y)$.

Acknowledgments

This work was supported in part by the National Science Foundation grants HRD-0734825 and DUE-0926721, by Grant 1 T36 GM078000-01 from the National Institutes of Health, and by Grant MSM 6198898701 from MŠMT of Czech Republic.

The author is thankful to all the participants of the Eighth International Conference on Computability and Complexity in Analysis CCA'2011 (Cape Town, South Africa, January 31 – February 4, 2011) for valuable discussions.

Литература

1. Bishop E. Foundations of constructive analysis. New York: McGraw-Hill, 1967.
2. Hoyrup M. Computable analysis and algorithmic randomness, // Abstracts of the Eighth International Conference on Computability and Complexity in Analysis CCA'2011, Cape Town, South Africa, January 31 – February 4, 2011. P. 15.
3. Finkelstein A.M., and Kreinovich V. Impossibility of hardly possible events: physical consequences // Abstracts of the 8th International Congress on Logic, Methodology, and Philosophy of Science, Moscow, 1987. V. 5, Pt. 2, P. 23-25.
4. Kreinovich V. Toward formalizing non-monotonic reasoning in physics: the use of Kolmogorov complexity // Revista Iberoamericana de Inteligencia Artificial 2009. V. 41, P. 4-20.
5. Kreinovich V. Negative results of computable analysis disappear if we restrict ourselves to random (or, more generally, typical) inputs // Abstracts of the Eighth International Conference on Computability and Complexity in Analysis CCA'2011, Cape Town, South Africa, January 31 – February 4, 2011. P. 15–16.
6. Kreinovich V. Under physics-motivated constraints, generally-non-algorithmic computational problems become algorithmically solvable // Abstracts of the Fourth International Workshop on Constraint Programming and Decision Making CoProD'11, El Paso, Texas, March 17, 2011.

7. Kreinovich, V., Finkelstein, A.M. Towards applying computational complexity to foundations of physics // Notes of Mathematical Seminars of St. Petersburg Department of Steklov Institute of Mathematics 2004. V. 316, P. 63-110; reprinted in Journal of Mathematical Sciences 2006. V. 134, N. 5, P. 2358-2382.
8. Kreinovich V., Kunin I.A. Kolmogorov complexity and chaotic phenomena // International Journal of Engineering Science 2003. V. 41, N. 3, P. 483-493.
9. Kreinovich V., Kunin I.A. Kolmogorov complexity: how a paradigm motivated by foundations of physics can be applied in robust control // In: A.L. Fradkov, A.N. Churilov, eds., Proceedings of the International Conference "Physics and Control" PhysCon'2003, Saint-Petersburg, Russia, August 20–22, 2003. P. 88-93.
10. Kreinovich V., Kunin I.A. Application of Kolmogorov complexity to advanced problems in mechanics // Proceedings of the Advanced Problems in Mechanics Conference APM'04, St. Petersburg, Russia, June 24–July 1, 2004. P. 241-245.
11. Kreinovich V., Longpré L., Koshelev M. Kolmogorov complexity, statistical regularization of inverse problems, and Birkhoff's formalization of beauty // In: A. Mohamad-Djafari, ed., Bayesian Inference for Inverse Problems, Proceedings of the SPIE/International Society for Optical Engineering, San Diego, California, 1998. V. 3459, P. 159-170.
12. Li M., Vitanyi P. An introduction to Kolmogorov complexity and its applications, Springer, 2008.
13. Pour-El M.B., Richards J.I. Computability in analysis and physics. Berlin:Springer, 1989.
14. Tikhonov A.N., Arsenin V.Y. Solutions of ill-posed problems. Washington, D.C.:W.H. Whinston & Sons, 1977.
15. Weihrauch, K. Computable analysis. Berlin:Springer-Verlag, 2000.