Negative Results of Computable Analysis Disappear If We Restrict Ourselves to Random (Or, More Generally, Typical) Inputs

Vladik Kreinovich
Department of Computer Science
University of Texas, El Paso, TX 79968, USA
vladik@utep.edu

1. 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 a measurement, and get a value $\widetilde{x}$.
- Measurements are never absolute accurate.
- We usually also know the upper bound Δ on the 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, e.g., 2^{-n} with $n = 0, 1, \ldots$
- So, we get a sequence of rational numbers r_n for which $|x - r_n| \le 2^{-n}$.
- Such sequences represent real numbers in computable analysis.

2. Known Negative Results

- No algorithm is possible that, given two numbers x and y, would check whether $x = y$.
- Similarly, we can define a computable function $f(x)$ from real numbers to real numbers as a mapping that:
  - given an integer n, a rational number x_m and its accuracy 2^{-m},
  - produces y_n which is 2^{-n}-close to all values $f(x)$ with $d(x, x_m) \le 2^{-m}$ (or nothing)

  so that for every x and for each desired accuracy n, there is an m for which a y_n is produced.
- We can similarly define a computable function $f(x)$ on a computable compact set K.
- No algorithm is possible that, given f, returns x s.t. $f(x) = \max\limits_{y \in K} f(y)$. (The max itself *is* computable.)

3. From the Physicists' Viewpoint, These Negative Results Seem Rather Theoretical

- In mathematics, if two numbers coincide up to 13 digits, they may still turn to be different.
- For example, they may be 1 and $1 + 10^{-100}$.
- In physics, if two quantities coincide up to a very high accuracy, it is a good indication that they are equal:
  - if an experimentally value is very close to the theoretical prediction,
  - this means that this theory is (triumphantly) true.
- This is how General Relativity was confirmed.
- This is how physicists realized that light is formed of electromagnetic waves: their speeds are very close.

4. How Physicists Argue

- In math, if two numbers coincide up to 13 digits, they may still turn to be different: e.g., 1 and $1 + 10^{-100}$.
- In physics, if two quantities coincide up to a very high accuracy, it is a good indication that they are equal.
- 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*).
- In physics, second order terms like $a \cdot \Delta x^2$ of the Taylor series can be ignored if Δx is small, since:
  - 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.

5. How to Formalize the Physicist's Intuition of Typical (Not Abnormal): Main Idea

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

6. How to Formalize the Physicist's Intuition of Typical (Not Abnormal): Resulting Definition

- **Definition.** *We thus 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 $\mathcal{T} \cap A_N = \emptyset$.*
- *Comment.* Of course, to make this definition precise,
  - we must restrict definability to a *subset* of properties,
  - so that the resulting notion of definability will be defined in ZFC itself.

7. Kolmogorov's Definition of Algorithmic Randomness

- *Kolmogorov:* 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,
    - 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.
- *What is a probability law:* a statement S which is true with probability 1: $P(S) = 1$.
- *Conclusion:* to prove that a sequence is not random, we must show that it does not satisfy one of these laws.

8. Kolmogorov's Definition of Algorithmic Randomness (cont-d)

- *Reminder:* a sequence s is not random if it does not satisfy one of the probability laws S.
- *Equivalent statement:* s is not random if $s \in C$ for a (definable) set C $(= -S)$ with $P(C) = 0$.
- *Resulting definition* (Kolmogorov, Martin-Löf): s is random if $s \notin C$ for all definable C with $P(C) = 0$.
- *Consistency proof:*
  - Every definable set C is defined by a finite sequence of symbols (its definition).
  - Since there are countably many sequences of symbols, there are countably many definable sets C.
  - So, the complement $-\mathcal{R}$ to the class $\mathcal{R}$ of all random sequences also has probability 0.

9. Towards a More Physically Adequate Versions of Kolmogorov Randomness

- *Problem:* the 1960s Kolmogorov's definition only explains why events with probability 0 do not happen.
- *What we need:* formalize the physicists' intuition that events with very small probability cannot happen.
- *Seemingly natural formalization:* there exists the "smallest possible probability" p_0 such that:
  - if the computed probability p of some event is larger than p_0, then this event can occur, while
  - if the computed probability p is $\le p_0$, the event cannot occur.
- *Example:* a fair coin falls heads 100 times with prob. 2^{-100}; it is impossible if $p_0 \ge 2^{-100}$.

10. The Above Formalization of Randomness is Not Always Adequate

- *Problem:* every sequence of heads and tails has exactly the same probability.
- *Corollary:* if we choose $p_0 \geq 2^{-100}$, we will thus exclude all sequences of 100 heads and tails.
- However, anyone can toss a coin 100 times.
- This proves that some such sequences are physically possible.
- *Similar situation:* Kyburg's lottery paradox:
  - in a big (e.g., state-wide) lottery, the probability of winning the Grand Prize is very small;
  - a reasonable person should not expect to win;
  - however, some people do win big prizes.

11. New Definition of Randomness

- *Example:* height:
  - if height is ≥ 6 ft, it is still normal;
  - if instead of 6 ft, we consider 6 ft 1 in, 6 ft 2 in, etc., then $\exists h_0$ s.t. everyone taller than h_0 is abnormal;
  - we are not sure what is h_0, but we are sure such h_0 exists.
- *General description:* on the universal set U, we have sets $A_1 \supseteq A_2 \supseteq \ldots \supseteq A_n \supseteq \ldots$ s.t. $P(\cap A_n) = 0$.
- *Example:* $A_1 =$ people w/height ≥ 6 ft, $A_2 =$ people w/height ≥ 6 ft 1 in, etc.
- A set $\mathcal{R} \subseteq U$ is called a *set of random elements* if

  $\forall$ definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ for all n and $P(\cap A_n) = 0$, $\exists N$ for which $A_N \cap \mathcal{R} = \emptyset$.

12. Coin Example

- Universal set $U = \{\mathrm{H}, \mathrm{T}\}^{\mathbb{N}}$
- 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 if a sequence $s \in \mathcal{R}$ is random and starts with N heads, it must consist of heads only.
- *In physical terms:* it means that

  a random sequence cannot start with N heads.

- This is exactly what we wanted to formalize.

13. Relation between Typical and Random

- A set $\mathcal{R} \subseteq U$ is called a *set of random elements* if

  $\forall$ definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ for all n and $P(\cap A_n) = 0$, $\exists N$ for which $A_N \cap \mathcal{R} = \emptyset$.

- A set $\mathcal{R} \subseteq U$ is called a *set of typical elements* if

  $\forall$ definable sequence of sets A_n for which $A_n \supseteq A_{n+1}$ for all n and $\cap A_n = \emptyset$, $\exists N$ for which $A_N \cap \mathcal{R} = \emptyset$.

- *Relation:* let $\mathcal{R}_K$ is the set of the elements random in the usual Komogorov-Martin-Löf sense. Then:
  - every set of random elements is also a set of typical elements (since if $\cap A_n = \emptyset$ then $P(A_n) \to 0$);
  - for every set of typical elements $\mathcal{T}$, the intersection $\mathcal{T} \cap \mathcal{R}_K$ is a set of random elements.

- If $P(\cap A_n) = 0$ then for $B_m \stackrel{\text{def}}{=} A_m - \cap A_n$, $B_m \supseteq B_{m+1}$, $\cap B_n = \emptyset$, so $\exists N\ (B_N \cap \mathcal{T} = \emptyset)$; and $(\cap A_n) \cap \mathcal{R}_K = \emptyset$.

14. Ill-Posed Problems: In Brief

- Main *objectives* of science:
  - *guaranteed* estimates for physical quantities;
  - *guaranteed* predictions for these quantities.
- *Problem:* estimation and prediction are ill-posed.
- *Example:*
  - measurement devices are inertial;
  - hence suppress high frequencies ω;
  - so $\varphi(x)$ and $\varphi(x)+\sin(\omega\cdot t)$ are indistinguishable.
- *Existing approaches:*
  - statistical regularization (filtering);
  - Tikhonov regularization (e.g., $|\dot{x}|\le\Delta$);
  - expert-based regularization.
- *Main problem:* no guarantee.

15. On "Not Abnormal" Solutions, Problems Become Well-Posed

- *State estimation – an ill-posed problem:*
  - *Measurement f:* state $s \in S \to$ observation $r = f(s) \in R$.
  - *In principle,* we can reconstruct $r \to s$: as $s = f^{-1}(r)$.
  - *Problem:* small changes in r can lead to huge changes in s (f^{-1} *not continuous*).
- *Theorem:*
  - Let S be a definably separable metric space.
  - Let $\mathcal{T}$ be a set of all not abnormal elements of S.
  - Let $f : S \to R$ be a continuous 1-1 function.
  - Then, the inverse mapping $f^{-1} : R \to S$ is *continuous* for every $r \in f(\mathcal{T})$.

16. Another Physically Interesting Consequence: Justification of Physical Induction

- *What is physical induction:* a property P is satisfied in the first N experiments, then it is satisfied always.
- *Comment:* N should be sufficiently large.
- *Theorem:* $\exists N$ s.t. if for a typical object o, P is satisfied in the first N experiments, then P is satisfied always.
- *Notation:* $s \stackrel{\text{def}}{=} s_1 s_2 \ldots$, where:
  - $s_i = T$ if P holds in the i-th experiment, and
  - $s_i = F$ if $\neg P$ holds in the i-th experiment.
- *Proof:* $A_n \stackrel{\text{def}}{=} \{o : s_1 = \ldots = s_n = T \,\&\, \exists m\,(s_m = F)\}$; then $A_n \supseteq A_{n+1}$ and $\cup A_n = \emptyset$ so $\exists N\,(A_N \cap \mathcal{T} = \emptyset)$.
- *Meaning* of $A_N \cap \mathcal{T} = \emptyset$: if $o \in \mathcal{T}$ and $s_1 = \ldots = s_N = T$, then $\neg\exists m\,(s_m = F)$, i.e., $\forall m\,(s_m = T)$.

17. When We Restrict Ourselves to Typical Elements, Algorithms Become Possible

- *New result:* 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.
- *Idea:* for $A_n = \{(x, y) : 0 < d(x, y) < 2^{-n}\}$, we have $A_n \supseteq A_{n+1}$ and $\cap A_n = \emptyset$, so $\exists N\ (A_N \cap \mathcal{T} = \emptyset)$.
- *Meaning:* if $(x, y) \in \mathcal{T}$, then $d(x, y) = 0$ (i.e., $x = y$) or $d(x, y) \geq 2^{-N}$.
- *Algorithm:* compute $d(x, y)$ with accuracy $2^{-(N+2)}$, i.e., compute d such that $|d(x, y) - d| \leq 2^{-(N+2)}$:
  - 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 $x = y$.

18. When We Restrict Ourselves to Typical Elements, Algorithms Become Possible (cont-d)

- There exists an algorithm that:
  - given a typical function $f(x)$ on a computable compact K,
  - computes a value x at which $f(x) = \max\limits_{y} f(y)$.
- 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 value x at which $f(x) = 0$.
- Moreover, we can compute 2^{-n}-approximations to the corresponding sets:

$$\{x : f(x) = \max_{y} f(y)\} \text{ and } \{x : f(x) = 0\}.$$

19. Proof: Main Idea

- To compute $R \stackrel{\text{def}}{=} \{x : f(x) = 0\}$ with accuracy $\varepsilon > 0$, take an $(\varepsilon/2)$-net $\{x_1, \ldots, x_n\} \subseteq K$.
- For each i, we can compute $\varepsilon' \in (\varepsilon/2, \varepsilon)$ for which $B_i \stackrel{\text{def}}{=} \{x : d(x, x_i) \le \varepsilon'\}$ is a computable compact set.
- Thus, we can compute $m_i \stackrel{\text{def}}{=} \min\{|f(x)| : x \in B_i\}$.
- As before, $\exists N\, \forall f \in \mathcal{T}\, \forall i\, (m_i = 0 \vee m_i \ge 2^{-N})$.
- 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 $d_H(R, \{x_i : m_i = 0\}) \le \varepsilon$.
- $m_i = 0 \Rightarrow \exists x\, (f(x) = 0\, \&\, d(x, x_i) < \varepsilon) \Rightarrow d(x_i, R) \le \varepsilon$.
- If $x \in R$, i.e., $f(x) = 0$, then $\exists i\, (d(x, x_i) \le \varepsilon/2)$ hence $m_i = 0$ and $x_i \in \{x_i : m_i = 0\}$.
- $f(x) = \max\limits_y f(y) \Leftrightarrow g(x) \stackrel{\text{def}}{=} f(x) - \max\limits_y f(y) = 0$.

20. Other Problems

- Is it possible to similarly compute the optimal *minimax* strategies, i.e., find x such that
$$\min_y f(x,y) = \max_z \min_y f(z,y)?$$
- Yes, this is the same as finding location of the maximum of a computable function $g(x) \stackrel{\text{def}}{=} \min_y f(x,y)$.
- It is possible to similarly compute *Pareto optimum* set:
  - we have several objective functions $f_1(x), \ldots, f_n(x)$;
  - we say that y is *better* than x if
$$\forall i\,(f_i(y) \ge f_i(x))\,\&\,\exists i\,(f_i(y) > f_i(x));$$
  - an alternative x is *Pareto-optimal* if no other alternative y is better than x.
- Is it possible to similarly compute the set of *local maxima* (*minima*)?

21. Acknowledgments

This work was supported in part:

- by National Science Foundation grants HRD-0734825, EAR-0225670, and DMS-0532645 and
- by Grant 1 T36 GM078000-01 from the National Institutes of Health.

22. Definable: Mathematical Comment

- *What is definable:*
  - 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 $\{x \mid P(x)\}$ is defined in $\mathcal{L}$.

  We will then call the set $\{x \mid P(x)\}$ *$\mathcal{L}$-definable.*

- *How to deal with definable sets:*
  - Our objective is to be able to make mathematical statements about $\mathcal{L}$-definable sets.
  - Thus, 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.

23. Consistency Proof

- *Statement:* $\forall \varepsilon > 0$, there exists a set $\mathcal{T}$ of typical elements for which $\underline{P}(\mathcal{T}) \geq 1 - \varepsilon$.
- There are countably many definable sequences $\{A_n\}$: $\{A_n^{(1)}\}$, $\{A_n^{(2)}\}$, ...
- For each k, $P\left(A_n^{(k)}\right) \to 0$ as $n \to \infty$.
- Hence, there exists N_k for which $P\left(A_{N_k}^{(k)}\right) \leq \varepsilon \cdot 2^{-k}$.
- We 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, $\underline{P}(\mathcal{T}) = 1 - \overline{P}\left(\bigcup\limits_{k=1}^{\infty} A_{N_k}^{(k)}\right) \geq 1 - \varepsilon$.

24. Proof of Well-Posedness

- *Known:* if a f is continuous and 1-1 on a compact, then f^{-1} is also continuous.
- *Reminder:* X is compact if and only if it is closed and for every ε, it has a finite ε-net.
- *Given:* S is definably separable.
- *Means:* $\exists$ def. $s_1, \ldots, s_n, \ldots$ everywhere dense in S.
- *Solution:* take $A_n \stackrel{\text{def}}{=} -\bigcup\limits_{i=1}^{n} B_\varepsilon(s_i)$.
- 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 $A_N = -\bigcup\limits_{i=1}^{N} B_\varepsilon(s_i)$, this means $\mathcal{T} \subseteq \bigcup\limits_{i=1}^{N} B_\varepsilon(s_i)$.
- Hence $\{s_1, \ldots, s_N\}$ is an ε-net for $\mathcal{T}$. Q.E.D.

25. Other Practical Use of Algorithmic Randomness: When to Stop an Iterative Algorithm

- *Situation* in numerical mathematics:
  - we often know an iterative process whose results x_k are known to converge to the desired solution x,
  - but we do not know when to stop to guarantee that

$$d_X(x_k, x) \le \varepsilon.$$

- *Heuristic approach:* stop when $d_X(x_k, x_{k+1}) \le \delta$ for some $\delta > 0$.
- *Example:* in physics, if 2nd order terms are small, we use the linear expression as an approximation.

26. When to Stop an Iterative Algorithm: Result

- Let $\{x_k\} \in S$, k be an integer, and $\varepsilon > 0$ a real number.
- We say that x_k is *ε-accurate* if $d_X(x_k, \lim x_p) \leq \varepsilon$.
- Let $d \geq 1$ be an integer.
- By a *stopping criterion*, we mean a function $c : X^d \to R_0^+$ that satisfies the following two properties:
  - If $\{x_k\} \in S$, then $c(x_k, \ldots, x_{k+d-1}) \to 0$.
  - If for some $\{x_n\} \in S$ and k, $c(x_k, \ldots, x_{k+d-1}) = 0$, then $x_k = \ldots = x_{k+d-1} = \lim x_p$.
- *Result:* Let c be a stopping criterion. Then, for every $\varepsilon > 0$, there exists a $\delta > 0$ such that
  - if $c(x_k, \ldots, x_{k+d-1}) \leq \delta$, and the sequence $\{x_n\}$ is not abnormal,
  - then x_k is ε-accurate.

27. References to Our Papers re Typical and Randomness

- Finkelstein, A.M., Kreinovich, V.: Impossibility of hardly possible events: physical consequences. Abstracts of the 8th International Congress on Logic, Methodology, and Philosophy of Science, Moscow, 1987, 5(2), 23–25 (1987)
- Kreinovich, V.: Toward formalizing non-monotonic reasoning in physics: the use of Kolmogorov complexity. Revista Iberoamericana de Inteligencia Artificial 41, 4–20 (2009)
- 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 316, 63–110 (2004); reprinted in Journal of Mathematical Sciences 134(5), 2358–2382 (2006)

28. References to Our Papers re Typical and Randomness (cont-d)

- Kreinovich, V., Kunin, I.A.: Kolmogorov complexity and chaotic phenomena. International Journal of Engineering Science 41(3), 483–493 (2003)
- Kreinovich, V., Kunin, I.A.: Kolmogorov complexity: how a paradigm motivated by foundations of physics can be applied in robust control. In: Fradkov, A.L., Churilov, A.N., eds. Proceedings of the International Conference "Physics and Control" PhysCon'2003, Saint-Petersburg, Russia, August 20–22, 2003, 88–93 (2003)
- 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, 241–245 (2004)

29. References to Our Papers re Typical and Randomness (cont-d)

- Kreinovich, V., Longpré, L., Koshelev, M.: Kolmogorov complexity, statistical regularization of inverse problems, and Birkhoff's formalization of beauty. In: Mohamad-Djafari, A., ed., Bayesian Inference for Inverse Problems, Proceedings of the SPIE/International Society for Optical Engineering, San Diego, California, 1998, 3459, 159–170 (1998)

30. References to Other Related Papers

- Li, M., Vitanyi, P.: An Introduction to Kolmogorov Complexity and Its Applications, Springer (2008)
- Pour-El, M.B., Richards, J.I.: Computability in Analysis and Physics, Springer, Berlin (1989)
- Weihrauch, K.: Computable Analysis, Springer-Verlag, Berlin (2000)