Need to Combine Interval and Probabilistic Uncertainty: What Needs to Be Computed, What Can Be Computed, What Can Be Feasibly Computed, and How Physics Can Help

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

http://www.cs.utep.edu/vladik

Part I

Need to Combine Interval and Probabilistic Uncertainty: Linearized Case

1. Need to Take Uncertainty Into Account When Processing Data

- In practice, we are often interested in a quantity y which is difficult to measure directly.
- *Examples:* distance to a star, amount of oil in the well, tomorrow's weather.
- *Solution:* find easier-to-measure quantities $x_1, \ldots, x_n$ related to y by a known dependence $y = f(x_1, \ldots, x_n)$.
- Then, we measure x_i and use measurement results $\widetilde{x}_i$ to compute an estimate $\widetilde{y} = f(\widetilde{x}_1, \ldots, \widetilde{x}_n)$.
- Measurements are never absolutely accurate, so even if the model f is exact, $\widetilde{x}_i \neq x_i$ leads to $\Delta y \stackrel{\text{def}}{=} \widetilde{y} - y \neq 0$.
- It is important to use information about measurement errors $\Delta x_i \stackrel{\text{def}}{=} \widetilde{x}_i - x_i$ to estimate the accuracy Δy.

2. We Often Have Imprecise Probabilities

- *Usual assumption:* we know the probabilities for Δx_i.
- To find them, we measure the same quantities:
  – with our measuring instrument (MI) and
  – with a much more accurate MI, with $\widetilde{x}_i^{\rm st} \approx x_i$.
- In two important cases, this does not work:
  – state-of-the art-measurements, and
  – measurements on the shop floor.
- Then, we have partial information about probabilities.
- Often, all we know is an upper bound $|\Delta x_i| \le \Delta_i$.
- Then, we only know that $x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]$ and

$$y \in [\underline{y}, \overline{y}] \stackrel{\rm def}{=} \{f(x_1, \ldots, x_n) : x_i \in [\widetilde{x}_i - \Delta_i, \widetilde{x}_i + \Delta_i]\}.$$

- Computing $[\underline{y}, \overline{y}]$ is known as *interval computation.*

3. Data Processing: Example

- *Example:*
  - we want to measure coordinates X_j of an object;
  - we measure the distance Y_i between this object and objects with accurately known coordinates $X_j^{(i)}$:

$$Y_i = \sqrt{\sum_{j=1}^{3} (X_j - X_j^{(i)})^2}.$$

- *General case:*
  - we know the results $\widetilde{Y}_i$ of measuring Y_i;
  - we want to estimate the desired quantities X_j.

4. Usually Linearization Is Possible

- In most practical situations, we know the approximate values $X_j^{(0)}$ of the desired quantities X_j.
- These approximation are usually reasonably good, in the sense that the difference $x_j \stackrel{\text{def}}{=} X_j - X_j^{(0)}$ are small.
- In terms of x_j, we have

$$Y_i = f(X_1^{(0)} + x_1, \ldots, X_n^{(0)} + x_n).$$

- We can safely ignore terms quadratic in x_j.
- Indeed, even if the estimation accuracy is 10% (0.1), its square is $1\% \ll 10\%$.
- We can thus expand the dependence of Y_i on x_j in Taylor series and keep only linear terms:

$$Y_i = Y_i^{(0)} + \sum_{j=1}^{n} a_{ij} \cdot x_j, \quad Y_i^{(0)} \stackrel{\text{def}}{=} f_i(X_1^{(0)}, \ldots, X_n^{(0)}), \quad a_{ij} \stackrel{\text{def}}{=} \frac{\partial f_i}{\partial X_j}.$$

5. Least Squares

- Thus, to find the unknowns x_j, we need to solve a system of approximate linear equations

$$\sum_{j=1}^{n} a_{ij} \cdot x_i \approx y_i, \text{ where } y_i \stackrel{\text{def}}{=} \widetilde{Y}_i - Y_i^{(0)}.$$

- Usually, it is assumed that each measurement error is:
  - normally distributed
  - with 0 mean (and known st. dev. σ_i).
- The distribution is indeed often normal:
  - the measurement error is a joint result of many independent factors,
  - and the distribution of the sum of many small independent errors is close to Gaussian;
  - this is known as the Central Limit Theorem.

6. Least Squares (cont-d)

- 0 mean also makes sense:
  - we calibrate the measuring instrument by comparing it with a more accurate,
  - so if there was a bias (non-zero mean), we delete it by re-calibrating the scale.
- It is also assumed that measurement errors of different measurements are independent.
- In this case, for each possible combination $x = (x_1, \ldots, x_n)$, the probability of observing $y_1, \ldots, y_m$ is:

$$\prod_{i=1}^{m} \left(\frac{1}{\sqrt{2\pi} \cdot \sigma_i} \cdot \exp\left(- \frac{\left(y_i - \sum_{j=1}^{n} a_{ij} \cdot x_j \right)^2}{2\sigma_i^2} \right) \right).$$

7. Least Squares (final)

- It is reasonable to select x_j for which this probability is the largest, i.e., equivalently, for which

$$\sum_{i=1}^{n} \frac{\left(y_i - \sum\limits_{j=1}^{n} a_{ij} \cdot x_j\right)^2}{\sigma_i^2} \to \min .$$

- The set S_γ of all possible combinations x is:

$$S_\gamma = \left\{ x : \sum_{i=1}^{n} \frac{\left(y_i - \sum\limits_{j=1}^{n} a_{ij} \cdot x_j\right)^2}{\sigma_i^2} \le \chi^2_{m-n,\gamma} \right\}.$$

- If $S = \emptyset$, this means that some measurements are outliers.

8. Need to Take into Account Systematic Error

- In the traditional approach, we assume that $y_i = \sum_{j=1}^{n} a_{ij} \cdot x_j + e_i$, where the meas. error e_i has 0 mean.
- Sometimes:
  - in addition to the random error $e_i^r \stackrel{\text{def}}{=} e_i - E[e_i]$ with 0 mean,
  - we also have a systematic error $e_i^s \stackrel{\text{def}}{=} E[e_i]$:

$$y_i = \sum_{j=1}^{n} a_{ij} \cdot x_j + e_i^r + e_i^s.$$

- Sometimes, we know the upper bound Δ_i: $|e_i^s| \le \Delta_i$.
- In other cases, we have different bounds $\Delta_i(p)$ corresponding to different degree of confidence p.
- What can we then say about x_j?

9. Combining Probabilistic and Interval Uncertainty: Main Idea

- If we knew the values e_i^s, then we would conclude that for $e_i^r = y_i - \sum_{j=1}^{n} a_{ij} \cdot x_j - e_i^s$, we have

$$\sum_{i=1}^{m} \frac{(e_i^r)^2}{\sigma_i^2} = \sum_{i=1}^{m} \frac{\left(y_i - \sum_{j=1}^{n} a_{ij} \cdot x_j - e_i^s\right)^2}{\sigma_i^2} \le \chi^2_{m-n,\gamma}.$$

- In practice, we do not know the values e_i^s, we only know that these values are in the interval $[-\Delta_i, \Delta_i]$.
- Thus, we know that the above inequality holds for some

$$e_i^s \in [-\Delta_i, \Delta_i].$$

10. Main Idea (cont-d)

- The above condition is equivalent to $v(x) \le \chi^2_{m-n,\gamma}$, where

$$v(x) \stackrel{\text{def}}{=} \min_{e_i^s \in [-\Delta_i, \Delta_i]} \sum_{i=1}^{m} \frac{\left(y_i - \sum_{j=1}^{n} a_{ij} \cdot x_j - e_i^s\right)^2}{\sigma_i^2}.$$

- So, the set S_γ of all combinations $X = (x_1, \ldots, x_n)$ which are possible with confidence $1 - \gamma$ is:

$$S_\gamma = \{x : v(x) \le \chi^2_{m-n,\gamma}\}.$$

- The range of possible values of x_j can be obtained by maximizing and minimizing x_j under the constraint

$$v(x) \le \chi^2_{m-n,\gamma}.$$

- In the fuzzy case, we have to repeat the computations for every p.

11. How to Check Consistency

- We want to make sure that the measurements are consistent – i.e., that there are no outliers.
- This means that we want to check that there exists some $x = (x_1, \ldots, x_n)$ for which $v(x) \le \chi^2_{m-n,\gamma}$.
- This condition is equivalent to

$$v \stackrel{\text{def}}{=} \min_x v(x) =$$

$$\min_x \min_{e_i^s \in [-\Delta_i, \Delta_i]} \sum_{i=1}^m \frac{\left(y_i - \sum_{j=1}^n a_{ij} \cdot x_j - e_i^s\right)^2}{\sigma_i^2} \le \chi^2_{m-n,\gamma}.$$

12. This Is Indeed a Generalization of Probabilistic and Interval Approaches

- In the case when $\Delta_i = 0$ for all i, i.e., when there is no interval uncertainty, we get the usual Least Squares.
- Vice versa, for very small σ_i, we get the case of pure interval uncertainty.
- In this case, the above formulas tend to the set of all the values for which $\left|y_i - \sum_{j=1}^{n} a_{ij} \cdot x_j\right| \le \Delta_i$.
- E.g., for m repeated measurements of the same quantity, we get the intersection of the corr. intervals.
- So, the new idea is indeed a generalization of the known probabilistic and interval approaches.

13. From Formulas to Computations

- The expression $\left(y_i - \sum_{j=1}^{n} a_{ij} \cdot x_j - e_i^s\right)^2$ is a convex function of x_j.
- The domain of possible values of $e^s = (e_1^s, \ldots, e_m^s)$ is also convex: it is a box

$$[-\Delta_1, \Delta_1] \times \ldots \times [-\Delta_m, \Delta_m].$$

- There exist efficient algorithms for computing minima of convex functions over convex domains.
- These algorithms also compute locations where these minima are attained.
- Thus, for every x, we can efficiently compute $v(x)$ and thus, efficiently check whether $v(x) \le \chi^2_{m-n,\gamma}$.
- Similarly, we can efficiently compute v and thus, check whether $v \le \chi^2_{m-n,\gamma}$ – i.e., whether we have outliers.

14. From Formulas to Computations (cont-d)

- The set S_γ is convex.
- We can approximate the set S_γ by:
  - taking a grid G,
  - checking, for each $x \in G$, whether $v(x) \le \chi^2_{m-n,\gamma}$, and
  - taking the convex hull of "possible" points.
- We can also efficiently find the minimum $\underline{x}_j$ of x_j over

$$x \in S_\gamma.$$

- By computing the min of $-x_j$, we can also find the maximum $\overline{x}_j$.

Part II
General Case: What Can Be Computed?

15. How Do We Describe Imprecise Probabilities?

- *Ultimate goal of most estimates:* to make decisions.
- *Known:* a rational decision-maker maximizes expected *utility* $E[u(y)]$.
- For smooth $u(y)$, $y \approx \widetilde{y}$ implies that

$$u(y) = u(\widetilde{y}) + (y - \widetilde{y}) \cdot u'(\widetilde{y}) + \frac{1}{2} \cdot (y - \widetilde{y})^2 \cdot u''(\widetilde{y}).$$

- So, to find $E[u(y)]$, we must know moments $E[(y - \widetilde{y})^k]$.
- Often, $u(y)$ abruptly changes: e.g., when pollution level exceeds y_0; then $E[u(y)] \sim F(y) \stackrel{\text{def}}{=} \text{Prob}(y \le y_0)$.
- From $F(y)$, we can estimate moments, so $F(y)$ is enough.
- Imprecise probabilities mean that we don't know $F(y)$, we only know bounds (*p-box*) $\underline{F}(y) \le F(y) \le \overline{F}(y)$.

16. What Is Computable?

- Computations with p-boxes are practically important.
- It is thus desirable to come up with efficient algorithms which are as general as possible.
- It is known that too general problems are often *not* computable.
- To avoid wasting time, it is therefore important to find out what *can* be computed.
- At first glance, this question sounds straightforward:
  - to describe a cdf, we can consider a computable function $F(x)$;
  - to describe a p-box, we consider a computable *function interval* $[\underline{F}(x), \overline{F}(x)]$.
- Often, we can do that, but we will show that sometimes, we need to go *beyond* function intervals.

17. Reminder: What Is Computable?

- A real number x corresponds to a value of a physical quantity.
- We can measure x with higher and higher accuracy.
- So, x is called *computable* if there is an algorithm, that, given k, produces a rational r_k s.t. $|x - r_k| \le 2^{-k}$.
- A *computable function* computes $f(x)$ from x.
- We can only use approximations to x.
- So, an algorithm for computing a function can, given k, request a 2^{-k}-approximation to x.
- Most usual functions are thus computable.
- *Exception:* step-function $f(x) = 0$ for $x < 0$ and $f(x) = 1$ for $x \ge 0$.
- Indeed, no matter how accurately we know $x \approx 0$, from $r_k = 0$, we cannot tell whether $x < 0$ or $x \ge 0$.

18. Consequences for Representing a cdf $F(x)$

- We would like to represent a general probability distribution by its cdf $F(x)$.
- From the purely mathematical viewpoint, this is indeed the most general representation.
- At first glance, it makes sense to consider computable functions $F(x)$.
- For many distributions, e.g., for Gaussian, $F(x)$ is computable.
- However, when $x = 0$ with probability 1, the cdf $F(x)$ is exactly the step-function.
- And we already know that the step-function is not computable.
- Thus, we need to find an alternative way to represent cdf's – beyond computable functions.

19. Back to the Drawing Board

- Each value $F(x)$ is the probability that $X \le x$.
- We cannot empirically find exact probabilities p.
- We can only estimate *frequencies* f based on a sample of size N.
- For large N, the difference $d \stackrel{\text{def}}{=} p - f$ is asymptotically normal, with $\mu = 0$ and $\sigma = \sqrt{\dfrac{p \cdot (1-p)}{N}}$.
- Situations when $|d - \mu| < 6\sigma$ are negligibly rare, so we conclude that $|f - p| \le 6\sigma$.
- For large N, we can get $6\sigma \le \delta$ for any accuracy $\delta > 0$.
- We get a sample $X_1, \ldots, X_N$.
- We don't know the exact values X_i, only measured values $\widetilde{X}_i$ s.t. $|\widetilde{X}_i - X_i| \le \varepsilon$ for some accuracy ε.
- So, what we have is a frequency $f = \text{Freq}(\widetilde{X}_i \le x)$.

20. Resulting Definition

- Here, $X_i \le x - \varepsilon \Rightarrow \widetilde{X}_i \le x \Rightarrow X_i \le x + \varepsilon$, so

$$\text{Freq}(X_i \le x - \varepsilon) \le f = \text{Freq}(\widetilde{X}_i \le x) \le \text{Freq}(X_i \le x + \varepsilon).$$

- Frequencies are δ-close to probabilities, so we arrive at the following:
- *For every x, $\varepsilon > 0$, and $\delta > 0$, we get a rational number f such that $F(x - \varepsilon) - \delta \le f \le F(x + \varepsilon) + \delta$.*
- This is how we define a computable cdf $F(x)$.
- In the computer, to describe a distribution on an interval $[\underline{T}, \overline{T}]$:
  - we select a grid $x_1 = \underline{T}$, $x_2 = \underline{T} + \varepsilon$, ..., and
  - we store the corr. frequencies f_i with accuracy δ.
- A class of possible distribution is represented, for each ε and δ, by a finite list of such approximations.

21. First Equivalent Definition

- *Original:* $\forall x\,\forall \varepsilon_{>0}\,\forall \delta_{>0}$, we get a rational f such that
$$F(x-\varepsilon)-\delta \le f \le F(x+\varepsilon)+\delta.$$
- *Equivalent:* $\forall x\,\forall \varepsilon_{>0}\,\forall \delta_{>0}$, we get a rational f which is δ-close to $F(x')$ for some x' such that $|x'-x| \le \varepsilon$.
- *Proof of equivalence:*
  - We know that $F(x+\varepsilon)-F(x+\varepsilon/3)\to 0$ as $\varepsilon \to 0$.
  - So, for $\varepsilon = 2^{-k}$, $k=1,2,\ldots$, we take f and f' s.t.
$$F(x+\varepsilon/3)-\delta/4 \le f \le F(x+(2/3)\cdot\varepsilon)+\delta/4$$
$$F(x+(2/3)\cdot\varepsilon)-\delta/4 \le f' \le F(x+\varepsilon)+\delta/4.$$
  - We stop when f and f' are sufficiently close:
$$|f-f'|\le \delta.$$
  - Thus, we get the desired f.

22. Second Equivalent Definition

- We start with pairs $(x_1, f_1), (x_2, f_2), \ldots$
- When $f_{i+1} - f_i > \delta$, we add intermediate pairs

$$(x_i, f_i + \delta), (x_i, f_i + 2\delta), \ldots, (x_i, f_{i+1}).$$

- The resulting set of pairs is (ε, δ)-close to the graph $\{(x, y) : F(x-0) \le y \le F(x)\}$ in Hausdorff metric d_H.
- (x, y) and (x', y') are (ε, δ)-*close* if $|x - x'| \le \varepsilon$ and $|y - y'| \le \delta$.
- The sets S and S' are (ε, δ)-close if:
  - for every $s \in S$, there is a (ε, δ)-close point $s' \in S'$;
  - for every $s' \in S'$, there is a (ε, δ)-close point $s \in S$.
- Compacts with metric d_H form a computable compact.
- So, $F(x)$ is a monotonic computable object in this compact.

23. What Can Be Computed: A Positive Result for the 1D Case

- *Reminder:* we are interested in $F(x)$ and $E_{F(x)}[u(x)]$ for smooth $u(x)$.
- *Reminder:* estimate for $F(x)$ is part of the definition.
- *Question:* computing $E_{F(x)}[u(x)]$ for smooth $u(x)$.
- *Our result:* there is an algorithm that:
  - given a computable cdf $F(x)$,
  - given a computable function $u(x)$, and
  - given accuracy $\delta > 0$,
  - computes $E_{F(x)}[u(x)]$ with accuracy δ.
- For computable classes $\mathcal{F}$ of cdfs, a similar algorithm computes the range of possible values

$$[\underline{u}, \overline{u}] \stackrel{\text{def}}{=} \{E_{F(x)}[u(x)] : F(x) \in \mathcal{F}\}.$$

24. Proof: Main Idea

- Computable functions are computably continuous: for every $\delta > 0$, we can compute $\varepsilon > 0$ s.t.
$$|x - x'| \le \varepsilon \Rightarrow |f(x) - f(x')| \le \delta.$$
- We select ε corr. to $\delta/4$, and take a grid with step $\varepsilon/4$.
- For each x_i, the value f_i is $(\delta/4)$-close to $F(x'_i)$ for some x'_i which is $(\varepsilon/4)$-close to x_i.
- The function $u(x)$ is $(\delta/2)$-close to a piece-wise constant function $u'(x) = u(x_i)$ for $x \in [x'_i, x'_{i+1}]$.
- Thus, $|E[u(x)] - E[u'(x)]| \le \delta/2$.
- Here, $E[u'(x)] = \sum\limits_i u(x_i) \cdot (F(x'_{i+1}) - F(x'_i))$.
- Here, $F(x'_i)$ is close to f_i and $F(x'_{i+1})$ is close to f_{i+1}.
- Thus, $E[u'(x)]$ (and hence, $E[u(x)]$) is computably close to a computable sum $\sum\limits_i u(x_i) \cdot (f_{i+1} - f_i)$.

25. What to Do in a Multi-D Case?

- For each $g(x)$, y, $\varepsilon > 0$, and $\delta > 0$, we can find a frequency f such that:

$$|P(g(x) \le y') - f| \le \varepsilon \text{ for some } y' \text{ s.t. } |y - y'| \le \delta.$$

- We select an ε-net $x_1, \ldots, x_n$ for X. Then,

$$X = \bigcup_i B_\varepsilon(x_i), \text{ where } B_\varepsilon(x) \stackrel{\text{def}}{=} \{x' : d(x, x') \le \varepsilon\}.$$

- We select f_1 which is close to $P(B_{\varepsilon'}(x_1))$ for all ε' from some interval $[\underline{\varepsilon}, \overline{\varepsilon}]$ which is close to ε.
- We then select f_2 which is close to $P(B_{\varepsilon'}(x_1) \cup B_{\varepsilon'}(x_2))$ for all ε' from some subinterval of $[\underline{\varepsilon}, \overline{\varepsilon}]$, etc.
- Then, we get approximations to probabilities of the sets $B_\varepsilon(x_i) - (B_\varepsilon(x_1) \cup \ldots \cup B_\varepsilon(x_{i-1}))$.
- This lets us compute the desired values $E[u(x)]$.

Part III
Taking Into Account that We Process Physical Data

26. Computations with Real Numbers: Reminder

- From the physical viewpoint, real numbers x describe values of different quantities.
- We get values of real numbers by measurements.
- Measurements are never 100% accurate, so after a measurement, we get an approximate value r_k of x.
- In principle, we can measure x with higher and higher accuracy.
- So, from the computational viewpoint, a real number is a sequence of rational numbers r_k for which, e.g.,

$$|x - r_k| \le 2^{-k}.$$

- By an algorithm processing real numbers, we mean an algorithm using r_k as an "oracle" (subroutine).
- This is how computations with real numbers are defined in *computable analysis*.

27. 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.)

28. 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.

29. 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.

30. How to Formalize the Physicist's Intuition of Physically Meaningful Values: 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 physically meaningful 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 $\mathcal{T} \cap A_N = \emptyset$.

31. How to Formalize the Physicist's Intuition: Resulting Definition

- **Definition.** *We thus say that $\mathcal{T}$ is a* set of physically meaningful elements *if:*
  - *for every definable decreasing sequence $\{A_n\}$ for which $\bigcap_n A_n = \emptyset$,*
  - *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.

32. Checking Equality of Real Numbers

- *Known:* equality of real numbers is undecidable.
- For physically meaningful real numbers, however, a deciding algorithm *is* possible:
  - *for every set* $\mathcal{T} \subseteq \mathbb{R}^2$ *which consists of physically meaningful pairs* (x, y) *of real numbers,*
  - *there exists an algorithm deciding whether* $x = y$.
- *Proof:* We can take $A_n = \{(x, y) : 0 < |x - y| < 2^{-n}\}$. The intersection of all these sets is empty.
- Hence, $\mathcal{T}$ has no elements from $\bigcap\limits_{n=1}^{N_A} A_n = A_{N_A}$.
- Thus, for each $(x, y) \in \mathcal{T}$, $x = y$ or $|x - y| \geq 2^{-N_A}$.
- We can detect this by taking $2^{-(N_A+3)}$-approximations x' and y' to x and y. Q.E.D.

33. Finding Roots

- In general, it is not possible, given a f-n $f(x)$ attaining negative and positive values, to compute its root.
- This becomes possible if we restrict ourselves to physically meaningful functions:
- *Let K be a computable compact.*
- *Let X be the set of all functions $f : K \to \mathbb{R}$ that attain 0 value somewhere on K. Then:*
  - *for every set $\mathcal{T} \subseteq X$ consisting of physically meaningful functions and for every $\varepsilon > 0$,*
  - *there is an algorithm that, given a f-n $f \in \mathcal{T}$, computes an ε-approximation to the set of roots*

$$R \stackrel{\text{def}}{=} \{x : f(x) = 0\}.$$

- In particular, we can compute an ε-approximation to one of the roots.

34. Optimization

- In general, it is not algorithmically possible to find x where $f(x)$ attains maximum.
- Let K be a computable compact. Let X be the set of all functions $f : K \to \mathbb{R}$. Then:
  - *for every set* $\mathcal{T} \subseteq X$ *consisting of physically meaningful functions and for every* $\varepsilon > 0$,
  - *there is an algorithm that, given a f-n* $f \in \mathcal{T}$, *computes an* ε*-approx. to* $S = \left\{x : f(x) = \max\limits_y f(y)\right\}$.
- In particular, we can compute an approximation to an individual $x \in S$.
- *Reduction to roots:* $f(x) = \max\limits_y f(y)$ iff $g(x) = 0$, where $g(x) \stackrel{\text{def}}{=} f(x) - \max\limits_y f(y)$.

35. Computing Fixed Points

- In general, it is not possible to compute all the fixed points of a given computable function $f(x)$.
- Let K be a computable compact. Let X be the set of all functions $f : K \to K$. Then:
  - *for every set* $\mathcal{T} \subseteq X$ *consisting of physically meaningful functions and for every* $\varepsilon > 0$,
  - *there is an algorithm that, given a f-n* $f \in \mathcal{T}$, *computes an* ε*-approximation to the set* $\{x : f(x) = x\}$.
- In particular, we can compute an approximation to an individual fixed point.
- *Reduction to roots:* $f(x) = x$ iff $g(x) = 0$, where $g(x) \stackrel{\text{def}}{=} d(f(x), x)$.

36. Computing Limits

- *In general:* it is not algorithmically possible to find a limit $\lim a_n$ of a convergent computable sequence.
- Let K be a computable compact. Let X be the set of all convergent sequences $a = \{a_n\}$, $a_n \in K$. Then:
  - *for every set $\mathcal{T} \subseteq X$ consisting of physically meaningful functions and for every $\varepsilon > 0$,*
  - *there exists an algorithm that, given a sequence $a \in \mathcal{T}$, computes its limit with accuracy ε.*
- *Use:* this enables us to compute limits of iterations and sums of Taylor series (frequent in physics).
- *Main idea:* for every $\varepsilon > 0$ there exists $\delta > 0$ such that when $|a_n - a_{n-1}| \leq \delta$, then $|a_n - \lim a_n| \leq \varepsilon$.
- *Intuitively:* we stop when two consequent iterations are close to each other.

Part IV

How to Take into Account that We Can Use Non-Standard Physical Phenomena to Process Data

37. Solving NP-Complete Problems Is Important

- In practice, we often need to find a solution that satisfies a given set of constraints.
- At a minimum, we need to check whether such a solution is possible.
- Once we have a candidate, we can feasibly check whether this candidate satisfies all the constraints.
- In theoretical computer science, "feasibly" is usually interpreted as computable in polynomial time.
- The class of all such problems is called NP.
- Example: satisfiability – checking whether a formula like $(v_1 \vee \neg v_2 \vee v_3) \,\&\, (v_4 \vee \neg v_2 \vee \neg v_5) \,\&\, \ldots$ can be true.
- Each problem from the class NP can be algorithmically solved by trying all possible candidates.

38. NP-Complete Problems (cont-d)

- For example, we can try all 2^n possible combinations of true-or-false values $v_1, \ldots, v_n$.
- For medium-size inputs, e.g., for $n \approx 300$, the resulting time 2^n is larger than the lifetime of the Universe.
- So, these exhaustive search algorithms are not practically feasible.
- It is not known whether problems from the class NP can be solved feasibly (i.e., in polynomial time).
- This is the famous open problem $\mathrm{P} \stackrel{?}{=} \mathrm{NP}$.
- We know that some problems are *NP-complete*: every problem from NP can be reduced to it.
- So, it is very important to be able to efficiently solve even one NP-hard problem.

39. Can Non-Standard Physics Speed Up the Solution of NP-Complete Problems?

- NP-complete means difficult to solve on computers based on the usual physical techniques.
- A natural question is: can the use of non-standard physics speed up the solution of these problems?
- This question has been analyzed for several specific physical theories, e.g.:
  - for quantum field theory,
  - for cosmological solutions with wormholes and/or casual anomalies.
- So, a scheme based on a theory may not work.

40. No Physical Theory Is Perfect

- If a speed-up is possible within a given theory, is this a satisfactory answer?
- In the history of physics,
  - always new observations appear
  - which are not fully consistent with the original theory.
- For example, Newton's physics was replaced by quantum and relativistic theories.
- Many physicists believe that every physical theory is approximate.
- For each theory T, inevitably new observations will surface which require a modification of T.
- Let us analyze how this idea affects computations.

41. No Physical Theory Is Perfect: How to Formalize This Idea

- *Statement:* for every theory, eventually there will be observations which violate this theory.
- To formalize this statement, we need to formalize what are *observations* and what is a *theory*.
- Most sensors already produce *observation* in the computer-readable form, as a sequence of 0s and 1s.
- Let ω_i be the bit result of an experiment whose description is i.
- Thus, all past and future observations form a (potentially) infinite sequence $\omega = \omega_1\omega_2\ldots$ of 0s and 1s.
- A physical *theory* may be very complex.
- All we care about is which sequences of observations ω are consistent with this theory and which are not.

42. What Is a Physical Theory?

- So, a physical theory T can be defined as the set of all sequences ω which are consistent with this theory.
- A physical theory must have at least one possible sequence of observations: $T \neq \emptyset$.
- A theory must be described by a finite sequence of symbols: the set T must be *definable.*
- How can we check that an infinite sequence $\omega = \omega_1\omega_2\ldots$ is consistent with the theory?
- The only way is check that for every n, the sequence $\omega_1 \ldots \omega_n$ is consistent with T; so:

$$\forall n\, \exists \omega^{(n)} \in T\, (\omega_1^{(n)} \ldots \omega_n^{(n)} = \omega_1 \ldots \omega_n) \Rightarrow \omega \in T.$$

- In mathematical terms, this means that T is *closed* in the Baire metric $d(\omega, \omega') \stackrel{\text{def}}{=} 2^{-N(\omega,\omega')}$, where

$$N(\omega, \omega') \stackrel{\text{def}}{=} \max\{k : \omega_1 \ldots \omega_k = \omega'_1 \ldots \omega'_k\}.$$

43. What Is a Physical Theory: Definition

- A theory must predict something new.
- So, for every sequence $\omega_1 \ldots \omega_n$ consistent with T, there is a continuation which does not belong to T.
- In mathematical terms, T is *nowhere dense.*
- *By a* physical theory, *we mean a non-empty closed nowhere dense definable set T.*
- *A sequence ω is* consistent with the no-perfect-theory principle *if it does not belong to any physical theory.*
- In precise terms, ω does not belong to the union of all definable closed nowhere dense set.
- There are countably many definable set, so this union is *meager* (= *Baire first category*).
- Thus, due to Baire Theorem, such sequences ω exist.

44. How to Represent Instances of an NP-Complete Problem

- For each NP-complete problem $\mathcal{P}$, its instances are sequences of symbols.
- In the computer, each such sequence is represented as a sequence of 0s and 1s.
- We can append 1 in front and interpret this sequence as a binary code of a natural number i.
- In principle, not all natural numbers i correspond to instances of a problem $\mathcal{P}$.
- We will denote the set of all natural numbers which correspond to such instances by $S_{\mathcal{P}}$.
- For each $i \in S_{\mathcal{P}}$, we denote the correct answer (true or false) to the i-th instance of the problem $\mathcal{P}$ by $s_{\mathcal{P},i}$.

45. What We Mean by Using Physical Observations in Computations

- In addition to performing computations, our computational device can:
  - produce a scheme i for an experiment, and then
  - use the result ω_i of this experiment in future computations.
- In other words, given an integer i, we can produce ω_i.
- In precise terms, the use of physical observations in computations means that use ω as an *oracle*.

46. Main Result

- A *ph-algorithm* $\mathcal{A}$ is an algorithm that uses an oracle ω consistent with the no-perfect-theory principle.
- The result of applying an algorithm $\mathcal{A}$ using ω to an input i will be denoted by $\mathcal{A}(\omega, i)$.
- We say that a feasible ph-algorithm $\mathcal{A}$ *solves almost all instances of an NP-complete problem* $\mathcal{P}$ if:

$$\forall \varepsilon_{>0}\, \forall n\, \exists N_{\geq n} \left(\frac{\#\{i \leq N : i \in S_{\mathcal{P}}\, \&\, \mathcal{A}(\omega, i) = s_{\mathcal{P},i}\}}{\#\{i \leq N : i \in S_{\mathcal{P}}\}} > 1 - \varepsilon \right).$$

- Restriction to sufficiently long inputs $N \geq n$ makes sense: for short inputs, we can do exhaustive search.
- **Theorem.** *For every NP-complete problem* $\mathcal{P}$*, there is a feasible ph-alg.* $\mathcal{A}$ *solving almost all instances of* $\mathcal{P}$.

47. This Result Is the Best Possible

- Our result is the best possible, in the sense that the use of physical observations cannot solve *all* instances:
- **Proposition.** *If P≠NP, then no feasible ph-algorithm $\mathcal{A}$ can solve all instances of $\mathcal{P}$.*
- Can we prove the result for *all* N starting with some N_0?
- We say that a feasible ph-algorithm $\mathcal{A}$ *δ-solves* $\mathcal{P}$ if

$$\exists N_0\,\forall N \geq N_0 \left(\frac{\#\{i \leq N : i \in S_{\mathcal{P}}\,\&\,\mathcal{A}(\omega, i) = s_{\mathcal{P},i}\}}{\#\{i \leq N : i \in S_{\mathcal{P}}\}} > \delta\right).$$

- **Proposition.** *For every NP-complete problem $\mathcal{P}$ and for every $\delta > 0$:*
  - *if there exists a feasible ph-algorithm $\mathcal{A}$ that δ-solves $\mathcal{P}$,*
  - *then there is a feasible algorithm $\mathcal{A}'$ that also δ-solves $\mathcal{P}$.*

Part V
Physical and Computational Consequences

48. 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:* $\forall \mathcal{T}\, \exists N$ s.t. if for $o \in \mathcal{T}$, $P(o)$ is satisfied in the first N experiments, then $P(o)$ is satisfied always.
- *Notation:* $s \stackrel{\text{def}}{=} s_1 s_2 \ldots$, where:
  - $s_i = T$ if $P(o)$ holds in the i-th experiment, and
  - $s_i = F$ if $\neg P(o)$ 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., $\forall m\, (s_m = T)$.

49. Ill-Posed Problem: Brief Reminder

- 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.

50. On Physically Meaningful 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 physically meaningful 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 $r \in f(\mathcal{T})$.

51. Everything Is Related: EPR Paradox

- Due to *Relativity Theory*, two spatially separated simultaneous events cannot influence each other.
- *Einstein, Podolsky, and Rosen* intended to show that in quantum physics, such influence is possible.
- *In formal terms*, let x and x' be measured values at these two events.
- *Independence* means that possible values of x do not depend on x', i.e., $\mathcal{T} = X \times X'$ for some X and X'.
- *Physical induction* implies that the pair (x, x') belongs to a set S of physically meaningful pairs.
- **Theorem.** *A set $\mathcal{T}$ os physically meaningful pairs cannot be represented as $X \times X'$.*
- Thus, everything *is* related – but we probably can't use this relation to pass information ($\mathcal{T}$ isn't computable).

52. 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.

53. When to Stop an Iterative Algorithm: Result

- Let $\{x_k\} \in \mathcal{T}$, k be an integer, and $\varepsilon > 0$ a real number.
- We say that x_k is ε-*accurate* if $d_X(x_k, \lim x_p) \le \varepsilon$.
- Let $d \ge 1$ be an integer.
- By a *stopping criterion*, we mean a function $c : X^d \to R_0^+$ that satisfies the following two properties:
  - If $\{x_k\} \in \mathcal{T}$, then $c(x_k, \ldots, x_{k+d-1}) \to 0$.
  - If for some $\{x_n\} \in \mathcal{T}$ 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 $c(x_k, \ldots, x_{k+d-1}) \le \delta$, and the sequence $\{x_n\}$ is physically meaningful,
  - then x_k is ε-accurate.

Part VI
Relation with Randomness

54. Towards Relation with Randomness

- If a sequence s is random, it satisfies all the probability laws such as the law of large numbers.
- 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.
- A probability law is a statement S which is true with probability 1: $P(S) = 1$.
- So, a sequence is random if it belongs to all definable sets of measure 1.
- A sequence belongs to a set of measure 1 iff it does not belong to its complement $C = -S$ with $P(C) = 0$.
- So, *a sequence is* random *if it does not belong to any definable set of measure 0.*

55. Randomness and Kolmogorov Complexity

- Different definabilities lead to different randomness.
- When definable means computable, randomness can be described in terms of Kolmogorov complexity

$$K(x) \stackrel{\text{def}}{=} \min\{\text{len}(p) : p \text{ generates } x\}.$$

- Crudely speaking, an infinite string $s = s_1 s_2 \ldots$ is random if, for some constant $C > 0$, we have

$$\forall n\,(K(s_1 \ldots s_n) \geq n - C).$$

- Indeed, if a sequence $s_1 \ldots s_n$ is truly random, then the only way to generate it is to explicitly print it:

$$\texttt{print}(s_1 \ldots s_n).$$

- In contrast, a sequence like 0101...01 generated by a short program is clearly not random.

56. From Kolmogorov-Martin-Löf Theoretical Randomness to a More Physical One

- The above 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, a kettle on a cold stove will not boil by itself – but the probability is non-zero.
- If a coin falls head 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.
- Indeed, then, for N for which $2^{-N} < p_0$, no sequence of N heads or tails would be possible at all.

57. From Kolmogorov-Martin-Löf Theoretical Randomness to a More Physical One (cont-d)

- We cannot have a universal threshold p_0 such that events with probability $\le p_0$ cannot happen.
- However, we know that:
  - for each decreasing ($A_n \supseteq A_{n+1}$) sequence of properties A_n with $\lim p(A_n) = 0$,
  - there exists an N above which a truly random sequence cannot belong to A_N.
- *Resulting definition:* 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 $\mathcal{R} \cap A_N = \emptyset$.

58. Random Sequences and Physically Meaningful Sequences

- Let $\mathcal{R}_K$ denote the set of all elements which are random in Kolmorogov-Martin-Löf sense. Then:
- *Every set of random elements consists of physically meaningful elements.*
- *For every set $\mathcal{T}$ of physically meaningful elements, the intersection $\mathcal{T} \cap \mathcal{R}_K$ is a set of random elements.*
- *Proof:* When A_n is definable, for $D_n \stackrel{\text{def}}{=} \bigcap\limits_{i=1}^{n} A_i - \bigcap\limits_{i=1}^{\infty} A_i$, we have $D_n \supseteq D_{n+1}$ and $\bigcap\limits_{n=1}^{\infty} D_n = \emptyset$, so $P(D_n) \to 0$.
- Therefore, there exists an N for which the set of random elements does not contain any elements from D_N.
- Thus, every set of random elements indeed consists of physically meaningful elements.

Part VII
Proofs

59. A Formal Definition of Definable Sets

- Let $\mathcal{L}$ be a theory.
- Let $P(x)$ be a formula from $\mathcal{L}$ for which the set $\{x \mid P(x)\}$ exists.
- We will then call the set $\{x \mid P(x)\}$ $\mathcal{L}$-*definable*.
- Crudely speaking, a set is $\mathcal{L}$-definable if we can explicitly *define* it in $\mathcal{L}$.
- All usual sets are definable: $\mathbb{N}$, $\mathbb{R}$, etc.
- Not every set is $\mathcal{L}$-definable:
  - every $\mathcal{L}$-definable set is uniquely determined by a text $P(x)$ in the language of set theory;
  - there are only countably many texts and therefore, there are only countably many $\mathcal{L}$-definable sets;
  - so, some sets of natural numbers are not definable.

60. How to Prove Results About Definable Sets

- Our objective is to be able to make mathematical statements about $\mathcal{L}$-definable sets. Therefore:
  - in addition to the theory $\mathcal{L}$,
  - we must have a stronger theory $\mathcal{M}$ in which the class of all $\mathcal{L}$-definable sets is a countable set.
- For every formula F from the theory $\mathcal{L}$, we denote its Gödel number by $\lfloor F \rfloor$.
- We say that a theory $\mathcal{M}$ is *stronger* than $\mathcal{L}$ if:
  - $\mathcal{M}$ contains all formulas, all axioms, and all deduction rules from $\mathcal{L}$, and
  - $\mathcal{M}$ contains a predicate $\text{def}(n, x)$ such that for every formula $P(x)$ from $\mathcal{L}$ with one free variable,

$$\mathcal{M} \vdash \forall y\, (\text{def}(\lfloor P(x) \rfloor, y) \leftrightarrow P(y)).$$

61. Existence of a Stronger Theory

- As $\mathcal{M}$, we take $\mathcal{L}$ plus all above equivalence formulas.
- Is $\mathcal{M}$ consistent?
- Due to compactness, we prove that for any $P_1(x), \ldots, P_m(x)$, $\mathcal{L}$ is consistent with the equivalences corr. to $P_i(x)$.
- Indeed, we can take

$$\text{def}(n, y) \leftrightarrow (n = \lfloor P_1(x) \rfloor \,\&\, P_1(y)) \vee \ldots \vee (n = \lfloor P_m(x) \rfloor \,\&\, P_m(y)).$$

- This formula is definable in $\mathcal{L}$ and satisfies all m equivalence properties.
- Thus, the existence of a stronger theory is proven.
- The notion of an $\mathcal{L}$-definable set can be expressed in $\mathcal{M}$: S is $\mathcal{L}$-definable iff $\exists n \in \mathbb{N}\, \forall y\, (\text{def}(n, y) \leftrightarrow y \in S)$.
- So, all statements involving definability become statements from the $\mathcal{M}$ itself, *not* from metalanguage.

62. Consistency Proof

- *Statement:* $\forall \varepsilon > 0$, there exists a set $\mathcal{T}$ 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, $\underline{P}(\mathcal{T}) = 1 - \overline{P}\left(\bigcup\limits_{k=1}^{\infty} A_{N_k}^{(k)}\right) \geq 1 - \varepsilon$.

63. Finding Roots: Proof

- To compute the set $R = \{x : f(x) = 0\}$ with accuracy $\varepsilon > 0$, let us 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.
- It is possible to algorithmically compute the minimum of a function on a computable compact set.
- Thus, we can compute $m_i \stackrel{\text{def}}{=} \min\{|f(x)| : x \in B_i\}$.
- Since $f \in T$, similarly to the previous proof, we can prove that $\exists N\, \forall f \in T\, \forall i\, (m_i = 0 \vee m_i \ge 2^{-N})$.
- Comp. m_i w/acc. $2^{-(N+2)}$, we check $m_i = 0$ or $m_i > 0$.
- Let's prove that $d_H(R, \{x_i : m_i = 0\}) \le \varepsilon$, i.e., that $\forall i\, (m_i = 0 \Rightarrow \exists x\, (f(x) = 0\, \&\, d(x, x_i) \le \varepsilon))$ and $\forall x\, (f(x) = 0 \Rightarrow \exists i\, (m_i = 0\, \&\, d(x, x_i) \le \varepsilon))$.

64. Finding Roots: Proof (cont-d)

- $m_i = 0$ means $\min\{|f(x)| : x \in B_i \stackrel{\text{def}}{=} B_{\varepsilon'}(x_i)\} = 0$.
- 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 $m_i = 0$.

65. Proof of Well-Posedness

- *Known:* if a f is continuous and 1-1 on a compact, then f^{-1} is also continuous.
- *Reminder:* S is compact if and only if it is closed and for every ε, it has a finite ε-net.
- *Given:* the set X is definably separable.
- *Means:* $\exists$ def. $s_1, \ldots, s_n, \ldots$ everywhere dense in X.
- *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.

66. Random Sequences and Physically Meaningful Sequences (proof cont-d)

- Let T consist of physically meaningful elements. Let us prove that $\mathcal{T} \cap \mathcal{R}_K$ is a set of random elements.
- If $A_n \supseteq A_{n+1}$ and $P\left(\bigcap\limits_{n=1}^{\infty} A_n\right) = 0$, then for $B_m \stackrel{\text{def}}{=} A_m - \bigcap\limits_{n=1}^{\infty} A_n$, we have $B_m \supseteq B_{m+1}$ and $\bigcap\limits_{n=1}^{\infty} B_n = \emptyset$.
- Thus, by definition of a set consisting of physically meaningful elements, we conclude that $B_N \cap T = \emptyset$.
- Since $P\left(\bigcap\limits_{n=1}^{\infty} A_n\right) = 0$, we also know that $\left(\bigcap\limits_{n=1}^{\infty} A_n\right) \cap \mathcal{R}_K = \emptyset$.
- Thus, $A_N = B_N \cup \left(\bigcap\limits_{n=1}^{\infty} A_n\right)$ has no common elements with the intersection $T \cap \mathcal{R}_K$. Q.E.D.

67. Using Non-Standard Physics: Proof of the Main Result

- As $\mathcal{A}$, given an instance i, we simply produce the result ω_i of the i-th experiment.
- Let us prove, by contradiction, that for every $\varepsilon > 0$ and for every n, there exists an integer $N \geq n$ for which

$$\#\{i \leq N : i \in S_{\mathcal{P}} \,\&\, \omega_i = s_{\mathcal{P},i}\} > (1-\varepsilon)\cdot\#\{i \leq N : i \in S_{\mathcal{P}}\}.$$

- The assumption that this property is not satisfied means that for some $\varepsilon > 0$ and for some integer n, we have

$$\forall N_{\geq n}\, \#\{i \leq N : i \in S_{\mathcal{P}} \,\&\, \omega_i = s_{\mathcal{P},i}\} \leq (1-\varepsilon)\cdot\#\{i \leq N : i \in S_{\mathcal{P}}\}.$$

- Let $T \stackrel{\text{def}}{=} \{x : \#\{i \leq N : i \in S_{\mathcal{P}} \,\&\, x_i = s_{\mathcal{P},i}\} \leq$

$$(1-\varepsilon)\cdot\#\{i \leq N : i \in S_{\mathcal{P}}\} \text{ for all } N \geq n\}.$$

- We will prove that this set T is a physical theory (in the sense of the above definition); then $\omega \notin T$.

68. Proof (cont-d)

- *Reminder:* $T = \{x : \#\{i \le N : i \in S_{\mathcal{P}} \,\&\, x_i = s_{\mathcal{P},i}\} \le$
$$(1-\varepsilon)\cdot \#\{i \le N : i \in S_{\mathcal{P}}\} \text{ for all } N \ge n\}.$$
- By definition, a physical theory is a set which is non-empty, closed, nowhere dense, and definable.
- Non-emptiness is easy: the sequence $x_i = \neg s_{\mathcal{P},i}$ for $i \in S_{\mathcal{P}}$ belongs to T.
- One can prove that T is closed, i.e., if $x^{(m)} \in T$ for which $x^{(m)} \to \omega$, then $x \in T$.
- Nowhere dense means that for every finite sequence $x_1 \ldots x_m$, there exists a continuation $x \notin T$.
- Indeed, for extension, we can take $x_i = s_{\mathcal{P},i}$ if $i \in S_{\mathcal{P}}$.
- Finally, we have an explicit definition of T, so T is definable.

69. Non-Standard Physics: Proof of First Proposition

- Let us assume that P≠NP; we want to prove that for every feasible ph-algorithm $\mathcal{A}$, it is not possible to have

$$\forall N\,(\#\{i \le N : i \in S_{\mathcal{P}} \,\&\, \mathcal{A}(\omega, i) = s_{\mathcal{P},i}\} = \#\{i \le N : i \in S_{\mathcal{P}}\}).$$

- Let us consider, for each feasible ph-algorithm $\mathcal{A}$,

$$T(\mathcal{A}) \stackrel{\text{def}}{=} \{x : \#\{i \le N : i \in S_{\mathcal{P}} \,\&\, \mathcal{A}(x, i) = s_{\mathcal{P},i}\} = \#\{i \le N : i \in S_{\mathcal{P}}\} \text{ for all } N\}.$$

- Similarly to the proof of the main result, we can show that this set $T(\mathcal{A})$ is closed and definable.
- To prove that $T(\mathcal{A})$ is nowhere dense, we extend $x_1 \ldots x_m$ by 0s; then $x \in T$ would mean P=NP.
- If $T(\mathcal{A}) \ne \emptyset$, then $T(\mathcal{A})$ is a theory, so $\omega \notin T(\mathcal{A})$.
- If $T(\mathcal{A}) = \emptyset$, this also means that $\mathcal{A}$ does not solve all instances of the problem $\mathcal{P}$ – no matter what ω we use.

70. Proof of Second Proposition

- Let us assume that no non-oracle feasible algorithm δ-solves the problem $\mathcal{P}$.
- Let's consider, for each N_0 and feasible ph-alg. $\mathcal{A}$,

$$T(\mathcal{A}, N_0) \stackrel{\text{def}}{=} \{x : \#\{i \le N : i \in S_{\mathcal{P}} \,\&\, \mathcal{A}(x, i) = s_{\mathcal{P},i}\} >$$
$$\delta \cdot \#\{i \le N : i \in S_{\mathcal{P}}\} \text{ for all } N \ge N_0\}.$$

- We want to prove that $\forall N_0\,(\omega \notin T(\mathcal{A}, N_0))$.
- Similarly to the proof of the Main Result, we can show that $T(\mathcal{A}, N_0)$ is closed and definable.
- To prove that $T(\mathcal{A}, N_0)$ is nowhere dense, we extend $x_1 \ldots x_m$ by 0s.
- If $T(\mathcal{A}, N_0) \neq \emptyset$, then $T(\mathcal{A}, N_0)$ is a theory hence $\omega \notin T(\mathcal{A}, N_0)$.
- If $T(\mathcal{A}, N_0) = \emptyset$, then also $\omega \notin T(\mathcal{A}, N_0)$.

71. References: Linearized Case

- Sun, L., Dbouk, H., Neumann, I., Schoen, S., Kreinovich, V.: Taking into account interval (and fuzzy) uncertainty can lead to more adequate statistical estimates, Proceedings of the 2017 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2017, Cancun, Mexico, October 16–18, 2017.
http://www.cs.utep.edu/vladik/2017/tr17-57a.pdf

72. References: What Can Be Computed in General?

- Kreinovich, V., Pownuk, A., Kosheleva, O.: Combining interval and probabilistic uncertainty: what is computable?, In: Pardalos, P., Zhigljavsky, A., Zilinskas, J. (eds.): Advances in Stochastic and Deterministic Global Optimization, Springer Verlag, Cham, Switzerland, 2016, p. 13–32. http://www.cs.utep.edu/vladik/2015/tr15-66a.pdf

73. References: Taking Into Account that We Process Physical Data and that that We Can Use Non-Standard Physical Phenomena to Process Data

- Kreinovich, V.: Negative results of computable analysis disappear if we restrict ourselves to random (or, more generally, typical) inputs, Mathematical Structures and Modeling 25, 100–103 (2012)
- Kosheleva, O., Zakharevich, M., Kreinovich, V.: If many physicists are right and no physical theory is perfect, then by using physical observations, we can feasibly solve almost all instances of each NP-complete problem, Mathematical Structures and Modeling 31, 4–17 (2014)

74. 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)

75. 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)

Need to Combine . . .
General Case: What . . .
Taking Into Account . . .
How to Take into . . .
Physical and . . .
Relation with . . .
Proofs
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 83 of 85
Go Back
Full Screen
Close
Quit

76. 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)

77. 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)