Intermediate Degrees are Needed for the World to be Cognizable: Towards a New Justification for Fuzzy Logic Ideas

Hung T. Nguyen[1], Vladik Kreinovich[2], J. Esteban Gamez[2], François Modave[2], and Olga Kosheleva[3]

[1] Department of Mathematical Sciences, New Mexico State University, Las Cruces, NM 88003, USA hunguyen@nmsu.edu
[2] Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968, USA, vladik@utep.edu, estebangamez@gmail.com, fmodave@utep.edu
[3] Department of Teacher Education, University of Texas at El Paso, El Paso, TX 79968, USA, olgak@utep.edu

Summary. Most traditional examples of fuzziness come from the analysis of commonsense reasoning. When we reason, we use words from natural language like "young", "well". In many practical situations, these words do not have a precise true-or-false meaning, they are fuzzy. One may therefore be left with an impression that fuzziness is a subjective characteristic, it is caused by the specific way our brains work.

However, the fact that that we are the result of billions of years of successful adjusting-to-the-environment evolution makes us conclude that everything about us humans is not accidental. In particular, the way we reason is not accidental, this way must reflect some real-life phenomena – otherwise, this feature of our reasoning would have been useless and would not have been abandoned long ago.

In other words, the fuzziness in our reasoning must have an objective explanation – in fuzziness of the real world.

In this paper, we first give examples of objective real-world fuzziness. After these example, we provide an explanation of this fuzziness – in terms of cognizability of the world.

1 Introduction

One of the main ideas behind Zadeh's fuzzy logic and its applications is that everything is a matter of degree.

We are often accustomed to think that every statement about a physical world is true or false:

- that an object is either a particle or a wave,
- that a person is either young or not,

- that a person is either well or ill,

but in reality, we sometimes encounter intermediate situations.

That everything is a matter of degree is a convincing empirical fact, but a natural question is: why? How can we explain this fact?

This is what we will try to do in this paper: come up with a theoretical explanation of this empirical fact.

Most traditional examples of fuzziness come from the analysis of commonsense reasoning. When we reason, we use words from natural language like "young", "well". In many practical situations, these words do not have a precise true-or-false meaning, they are fuzzy. One may therefore be left with an impression that fuzziness is a subjective characteristic, it is caused by the specific way our brains work.

However, the fact that that we are the result of billions of years of successful adjusting-to-the-environment evolution makes us conclude that everything about us humans is not accidental. In particular, the way we reason is not accidental, this way must reflect some real-life phenomena – otherwise, this feature of our reasoning would have been useless and would not have been abandoned long ago.

In other words, the fuzziness in our reasoning must have an objective explanation – in fuzziness of the real world.

In this paper, we first give examples of objective real-world fuzziness. After these example, we provide an explanation of this fuzziness – in terms of cognizability of the world.

Comment. Some of our results first appeared in the conference papers [4, 16].

2 Examples of Objective "Fuzziness"

Fractals. The notion of dimension has existed for centuries. Already the ancient researchers made a clear distinction between 0-dimensional objects (points), 1-dimensional objects (lines), 2-dimensional objects (surfaces), 3-dimensional objects (bodies), etc.

In all these examples, dimension is a natural number: 0, 1, 2, 3, ...

Since the 19th century, mathematicians have provided a mathematical extension of the notion of dimension that allowed them to classify some weird mathematical sets as being of fractional (non-integer) dimension, but for a long time, these weird sets remained anomalies.

In the 1970s, B. Mandlebrot noticed that actually, many real-life objects have fractional dimension, ranging from the shoreline of England to the shape of the clouds and mountains to noises in electric circuits (to social phenomena such as stock prices). He called such sets of fractional (non-integer) dimension *fractals*; see, e.g., [11, 12, 13].

It is now clear that fractals play an important role in nature. So, what we originally thought of as an integer-valued variable turned out to be real-valued.

Quantum physics. Until the 19th century, physical phenomena were described by classical physics. In classical physics, some variables are continuous, some are discrete.

For example, the coordinates and velocities of particles usually take continuous values. However, if we are interested in stable states or periodic trajectories, we often end up with a discrete set of stable states.

This discreteness underlies most engineering implementations of computers: to represent 0 or 1, we select an object with 2 possible states, and use one of these states to represent 0 and another to represent 1.

In the 20th century, however, it turned out that a more adequate description of the physical world comes from *quantum physics*. One of the peculiar features of quantum physics is the so-called *superposition principle* (see, e.g., [2]) according to which with every two states $\langle 0|$ and $\langle 1|$, it is also possible to have "intermediate" states (superpositions) $c_0 \cdot \langle 0| + c_1 \cdot \langle 1|$ for all complex values c_0 and c_1 for which $|c_0|^2 + |c_1|^2 = 1$.

So, what we originally thought of as an integer-valued variable turned out to be real-valued.

Comment. It is worth mentioning that these quantum combinations of 0 and 1 states are not only happening in real life, but, as it was discovered in the 1990s, their use can drastically speed up computations. For example:

- we can search in an unsorted list of n elements in time $\sqrt{n}$ – which is much faster than the time n which is needed on non-quantum computers [6, 7, 18];
- we can factor a large integer in time which does not exceed a polynomial of the length of this integer – and thus, we can break most existing cryptographic codes like widely used RSA codes which are based on the difficulty of such a factorization on non-quantum computers [18, 21, 22].

These techniques form the basis of *quantum computing*; see, e.g., [18].

Fractional charges of quarks. In the late 19th century and early 20th century, it was experimentally confirmed that seemingly continuous matter is actually discrete: it consists of molecules, molecules consist of atoms, and atoms consist of elementary particles.

A part of this confirmation came from an experimental discovery that all electric charges are proportional to a single charge – which was later revealed to be equal to the charge of an electron.

Based on this proportionality, physicists concluded that many observed elementary particles ranging from (relatively) stables particles such as protons and neutrons to numerous unstable ones – like many mesons and baryons discovered in super-collides and in cosmic rays – cannot be further decomposed into "more elementary" objects.

In the 1960s, M. Gell-Mann [2, 5, 20] discovered that if we allow particles with fractional electronic charge, then we can describe protons, neutrons,

mesons, and baryons as composed of 3 (now more) even more elementary particles called *quarks*. At first, quarks were often viewed as purely mathematical constructions, but experiments with particle-particle collisions revealed that, within a proton, there are three areas (*partons*) where the reflection seems to be the largest – in perfect accordance with the fact that in the quark model, a proton consists of exactly three quarks.

So, what we originally thought of as an integer-valued variable turned out to be real-valued.

There exist other examples of objective "fuzziness". In physics, there are many other examples when what we originally thought of as an integer-valued variable turned out to be real-valued. In this paper, we just described the most well known ones.

3 Our Explanation of Why Physical Quantities Originally Thought to Be Integer-Valued Turned out to Be Real-Valued: Main Idea

In philosophical terms, what we are doing is "cognizing" the world, i.e., understanding how it works and trying to predict consequences of different actions – so that we will be able to select an action which is the most beneficial for us.

Of course, our knowledge is far from complete, there are many real-world phenomena which we have not cognized yet – and many philosophers believe that some of these phenomena are not cognizable at all.

If a phenomenon is not cognizable, there is nothing we can do about it. What we are interested in is phenomena which are cognizable. This is what we will base our explanation on – that in such phenomena, it is reasonable to expect continuous-valued variables, i.e., to expect that properties originally thought to be discrete are actually matters of degree.

4 First Explanation: Gödel's Theorem vs. Tarski's Algorithm

Gödel's theorem: a brief reminder. Our first explanation of "objective fuzziness" is based on the historically first result in which something was actually proven to be not cognizable – the well-known 1931 Gödel's theorem; see, e.g., [3].

This theorem can be formulated in terms of arithmetic. Specifically, we have variables which run over natural numbers 0, 1, 2, ... A term is anything that can be obtained from these variables and natural-valued constants by using addition and multiplication, e.g., $2 \cdot x \cdot y + 3 \cdot z$ (subtraction is also allowed).

Elementary formulas are defined as expressions of the type $t = t'$, $t < t'$, $t > t'$, $t \le t'$, $t \ge t'$, and $t \ne t'$ for some terms t and t'. Examples are $2 \cdot x \cdot y + 3 \cdot z = 0$ or $x < y + z$.

Finally, a formula is anything which is obtained from elementary formulas by using logical connectives "and" ($\&$), "or" ($\vee$), "implies" ($\rightarrow$), "not" ($\neg$), and quantifiers "for all x" ($\forall x$) and "there exists x" ($\exists x$). Example:

$$\forall x\, \forall y (x < y \rightarrow \exists z (y = x + y)).$$

Many statements about the physical world can be formulated in terms of such formulas. Our objective is therefore to find out whether a given formula is true or false.

Gödel's theorem states that no algorithm is possible that would, given a formula, check whether this formula is true or false. In other words, if we allow variables with discrete values, then it is not possible to have an algorithm which would solve all the problems.

Tarksi's result. In the 1940s, another well-known logician, Alfred Tarski, raised an interesting question: what if we only allow continuous variables? In other words, what if we consider the same formulas as Gödel considered, but we change their interpretation: now every variable can take arbitrary real values. It turns out that in this case, it is possible to have an algorithm that, given a formula, checks whether this formula is true or false. [23].

Conclusion. Thus, in a cognizable situations, we cannot have variables which only take discrete values – these variables must be able to take arbitrary real values.

Comment. It is worth mentioning that the original Tarski's algorithm required an unrealistically large amount of computation time; however, later, faster, practically useful algorithms have been invented; see, e.g., [1, 14].

5 Second Explanation: Efficient Algorithms vs. NP-Hardness

Not all algorithms are practical. Our first explanation of continuity (and "fuzziness") was that with the discrete variables, we cannot have a deciding algorithm, but with continuous variables, we can.

The existence of an algorithm is necessary for cognition, but not sufficient. It is well known that some theoretical algorithms are not practical at all. For example, if an algorithm requires an exponential number of computational steps 2^n on an input of size n, this means that for inputs of a reasonable size $n \approx 300 - 400$, the required computation time exceeds the lifetime of the Universe.

Feasible vs. non-feasible algorithms. There is still no perfect formalization of this difference between "practical" (*feasible*) and impractical (*non-feasible*) algorithms. Usually:

- algorithms for which the computation time $t_A(x)$ is bounded by some polynomial $P(n)$ of the length $n = \text{len}(x)$ of the input (e.g., linear-time, quadratic-time, etc.) are practically useful, while
- for practically useless algorithms, the computation time grows with the size of the input much faster than a polynomial.

In view of this empirical fact, in theoretical computer science, algorithms are usually considered *feasible* if their running time is bounded by a polynomial of n. The class of problems which can be solved in polynomial time is usually denoted by P; see, e.g., [19].

Notion of NP-hardness. Not all practically useful problems can be solved in polynomial time. To describe such problems, researchers have defined several more general classes of problems. One of the most well known classes is the class NP. By definition, this class consists of all the problems which can be solved in *non-deterministic* polynomial time – meaning that if we have a guess, we can check, in polynomial time, whether this guess is a solution to our problem.

Most computer scientists believe that NP≠P, i.e., that some problems from the class NP cannot be solved in polynomial time. However, this inequality has not been proven, it is still an open problem. What *is* known is that some problems are *NP-hard*, i.e., any problem from the class NP can be reduced to each of these problems in polynomial time. One of such NP-hard problems is the problem SAT of propositional satisfiability: given a propositional formula F, i.e., a formula obtained from Boolean (yes-no) variables $x_1, \ldots, x_n$ by using $\&$, $\vee$, and $\neg$, check whether there exist values $x_1, \ldots, x_n$ which make this formula true.

NP-hardness of SAT means that if NP≠P (i.e., if at least one problem from the class NP cannot be solved in polynomial time), then SAT also cannot be solved in polynomial time. In other words, SAT is the hardest of the problems from this class.

It is known that all the problems from the class NP can be solved in exponential time. Indeed, for a problem of size n, there are $\le a^n$ possible guesses, where a is the size of the corresponding alphabet, so we can simply try all these guesses one by one.

Example: systems of linear equations. One of the simplest-to-solve numerical problems is the solution to a system of linear equations

$$a_{11} \cdot x_1 + \ldots + a_{1n} \cdot x_n = b_1;$$

$$\ldots$$

$$a_{m1} \cdot x_1 + \ldots + a_{mn} \cdot x_n = b_m.$$

In the situation when all the unknowns x_i can take arbitrary real values, there exist efficient algorithms for solving such systems of equations – even the well-known Gauss elimination method, while not the fastest, it still feasible.

However, as soon as we restrict ourselves to discrete (e.g., integer-valued) variables x_i, the solution of such a system becomes an NP-hard problem [19].

Conclusion. So, we end up with the same conclusion: that in a cognizable situations, we cannot have variables which only take discrete values – these variables must be able to take arbitrary real values.

6 Case Study: Selecting the Most Representative Sample

Introduction to the problem. In many practical situations, it is desirable to find the statistical analysis of a certain population, but this population is so large that it is not practically possible to analyze every individual element from this population. In this case, we select a *sample* (subset) of the population, perform a statistical analysis on this sample, and use these results as an approximation to the desired statistical characteristics of the population as a whole.

For example, this is how polls work: instead of asking the opinion of all the people, pollsters ask a representative sample, and use the opinion of this sample as an approximation to the opinion of the whole population.

The more "representative" the sample, the larger our confidence that the statistical results obtained by using this sample are indeed a good approximation to the desired population statistics. Typically, we gauge the representativeness of a sample by how well its statistical characteristics reflect the statistical characteristics of the entire population. For example, in the sample of human voters, it is reasonable to require that in the selected sample, the average age, the average income, and other characteristics are the same as in the population in a whole. Of course, the representativeness of averages is not enough: e.g., the voting patterns of people whose salary is exactly the national average are not necessarily a good representation of how people will work on average. For that, we need the sample to include both poorer and reacher people – i.e., in general, to be representative not only in terms of averages but also in terms of, e.g., standard deviations (i.e., equivalently, in terms of variances).

In practice, many techniques are used to design a representative sample; see, e.g., [10]. In this section, we show that the corresponding exact optimization problem is computationally difficult (NP-hard).

How is this result related to fuzzy techniques? The main idea behind fuzzy techniques is that they formalize expert knowledge expressed by words from natural language.

In this section, we show that if we do not use this knowledge, i.e., if we only use the data, then selecting the most representative sample becomes a computationally difficult (NP-hard) problem. Thus, the need to find such samples in reasonable time justifies the use of fuzzy techniques.

Comment. Similar results are known: for example, it is known that a similar problem of maximizing diversity is NP-hard; see, e.g., [9].

Towards formulation of the problem in exact terms. Let us assume that we have a population consisting of N objects. For each of N objects, we know the values of k characteristics x_1, x_2, ..., x_k. The value of the first characteristic x_1 for i-th object will be denoted by $x_{1,i}$, the value of the second characteristic x_2 for the i-th object will be denoted by $x_{2,i}$, ..., and finally, the value of the characteristic x_k for the i-th object will be denoted by $x_{k,i}$. As a result, we arrive at the following formal definition:

Definition 1. *By a* population*, we mean a tuple*

$$p \stackrel{\text{def}}{=} \langle N, k, \{x_{j,i}\}\rangle,$$

where:

- *N is an integer; this integer will be called the* population size*;*
- *k is an integer; this integer is called the* number of characteristics*;*
- *$x_{j,i}$ ($1 \le j \le k, 1 \le i \le N$) are real numbers; the real number $x_{j,i}$ will be called the* value *of the j-th characteristic for the i-th object.*

Based on these known values, we can compute the population means

$$E_1 = \frac{1}{N} \cdot \sum_{i=1}^{N} x_{1,i}, \quad E_2 = \frac{1}{N} \cdot \sum_{i=1}^{N} x_{2,i}, \quad \ldots,$$

and the population variances

$$V_1 = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{1,i} - E_1)^2, \quad V_2 = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{2,i} - E_2)^2, \quad \ldots$$

We can also compute higher order central moments.

Definition 2. *Let $p = \langle N, k, \{x_{j,i}\}\rangle$ be a population, and let j be an integer from 1 to k.*

- *By the* population mean *E_j of the j-th characteristic, we mean the value*

$$E_j = \frac{1}{N} \cdot \sum_{i=1}^{N} x_{j,i}.$$

- *By the* population variance *V_j of the j-th characteristic, we mean the value*

$$V_j = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{j,i} - E_j)^2.$$

- *For every integer $d \ge 1$, by the even order* population central moment *$M_j^{(2d)}$* of order *$2d$ of the j-th characteristic, we mean the value*

$$M_j^{(2d)} = \frac{1}{N} \cdot \sum_{i=1}^{N} (x_{j,i} - E_j)^{2d}.$$

Comment. In particular, the population central moment $M_j^{(2)}$ of order 2 (corresponding to $d = 1$) is simply the population variance.

In addition to the values $x_{1,i}$, $x_{2,i}$, ..., we are given a size $n < N$ of the desirable sample. For each sample $I = \{i_1, \ldots, i_n\} \subseteq \{1, 2, \ldots, N\}$ of size n, we can compute the sample means

$$E_1(I) = \frac{1}{n} \sum_{i \in I} x_{1,i}, \quad E_2(I) = \frac{1}{n} \sum_{i \in I} x_{2,i}, \quad \ldots$$

and the sample variances

$$V_1(I) = \frac{1}{n} \sum_{i \in I} (x_{1,i} - E_1(I))^2, \quad V_2(I) = \frac{1}{n} \sum_{i \in I} (x_{2,i} - E_2(I))^2, \ldots$$

Definition 3. *Let N be a population size.*

- *By a* sample*, we mean a non-empty subset*

$$I \subseteq \{1, 2, \ldots, N\}.$$

- *For every sample I, by its* size*, we mean the number of elements in I.*

Definition 4. *Let $p = \langle N, k, \{x_{j,i}\}\rangle$ be a population, let I be a sample of size n, and let j be an integer from 1 to k.*

- *By the* sample mean $E_j(I)$ *of the j-th characteristic, we mean the value*

$$E_j(I) = \frac{1}{n} \cdot \sum_{i \in I} x_{j,i}.$$

- *By the* sample variance $V_j(I)$ *of the j-th characteristic, we mean the value*

$$V_j(I) = \frac{1}{n} \cdot \sum_{i \in I} (x_{j,i} - E_j(I))^2.$$

- *For every $d \geq 1$, by the* sample central moment $M_j^{(2d)}(I)$ of order $2d$ *of the j-th characteristic, we mean the value*

$$M_j^{(2d)}(I) = \frac{1}{n} \cdot \sum_{i \in I} (x_{j,i} - E_j(I))^{2d}.$$

Comment. Similarly to the population case, the sample central moment $M_j^{(2)}$ of order 2 (corresponding to $d = 1$) is simply the sample variance.

We want to select the *most representative* sample, i.e., the sample I for which the sample statistics $E_1(I), E_2(I), \ldots, V_1(I), V_2(I), \ldots$ are the closest to the population statistics $E_1, E_2, \ldots, V_1, V_2, \ldots$

Definition 5. *Let $p = \langle N, k, \{x_{j,i}\}\rangle$ be a population.*

- *By an* E-statistics tuple *corresponding to p, we mean a tuple*

$$t^{(1)} \stackrel{\text{def}}{=} (E_1, \ldots, E_k).$$

- *By an* (E, V)-statistics tuple *corresponding to p, we mean a tuple*

$$t^{(2)} \stackrel{\text{def}}{=} (E_1, \ldots, E_k, V_1, \ldots, V_k).$$

- *For every integer $d \geq 1$, by a* statistics tuple of order $2d$ *corresponding to p, we mean a tuple*

$$t^{(2d)} \stackrel{\text{def}}{=} (E_1, \ldots, E_k, M_1^{(2)}, \ldots, M_k^{(2)}, M_1^{(4)}, \ldots, M_k^{(4)}, \ldots, M_1^{(2d)}, \ldots, M_k^{(2d)}).$$

Comment. In particular, the statistics tuple of order 2 is simply the (E, V)-statistics tuple.

Definition 6. *Let $p = \langle N, k, \{x_{j,i}\}\rangle$ be a population, and let I be a sample.*

- *By an* E-statistics tuple *corresponding to I, we mean a tuple*

$$t^{(1)}(I) \stackrel{\text{def}}{=} (E_1(I), \ldots, E_k(I)).$$

- *By an* (E, V)-statistics tuple *corresponding to I, we mean a tuple*

$$t^{(2)}(I) \stackrel{\text{def}}{=} (E_1(I), \ldots, E_k(I), V_1(I), \ldots, V_k(I)).$$

- *For every integer $d \geq 2$, by a* statistics tuple of order $2d$ *corresponding to I, we mean a tuple*

$$t^{(2d)}(I) \stackrel{\text{def}}{=} (E_1(I), \ldots, E_k(I), M_1^{(2)}(I), \ldots, M_k^{(2)}(I),$$
$$M_1^{(4)}(I), \ldots, M_k^{(4)}(I) \ldots, M_1^{(2d)}(I), \ldots, M_k^{(2d)}(I)).$$

Comment. In particular, the statistics tuple of order 2 corresponding to a sample I is simply the (E, V)-statistics tuple corresponding to this same tuple.

We will show that no matter how we define closeness, this problem is NP-hard (computationally difficult).

Let us describe the problem in precise terms. To describe which tuple

$$t(I) \stackrel{\text{def}}{=} (E_1(I), E_2(I), \ldots, V_1(I), V_2(I), \ldots)$$

is the closest to the original statistics tuple

$$t \stackrel{\text{def}}{=} (E_1, E_2, \ldots, V_1, V_2, \ldots),$$

we need to fix a *distance function* $\rho(t(I), t)$ describing how distant are the two given tuples. Similarly to the usual distance, we would like this distance function to be equal to 0 when the tuples coincide and to be positive if when the tuples are different. So, we arrive at the following definitions.

Definition 7. *By a* distance function*, we mean a mapping ρ that maps every two real-valued tuples t and t' of the same size into a real value $\rho(t, t')$ in such a way that $\rho(t, t) = 0$ for all tuples t and $\rho(t, t') > 0$ for all $t \neq t'$.*

As an example, we can take Euclidean metric between the tuples $t = (t_1, t_2, \ldots)$ and $t' = (t'_1, t'_2, \ldots)$:

$$\rho(t, t') = \sqrt{\sum_j (t_j - t'_j)^2}.$$

Now, we are ready to formulate the problem.

Definition 8. *Let ρ be a distance function. By a* E-sample selection problem *corresponding to ρ, we mean the following problem. We are given:*

- *a population $p = \langle N, k, \{x_{j,i}\}\rangle$, and*
- *an integer $n < N$.*

Among all samples $I \subseteq \{1, \ldots, N\}$ of size n, we must find the sample I for which the distance $\rho(t^{(1)}(I), t^{(1)})$ between the corresponding E-statistical tuples is the smallest possible.

Definition 9. *Let ρ be a distance function. By a* (E, V)-sample selection problem *corresponding to ρ, we mean the following problem. We are given:*

- *a population $p = \langle N, k, \{x_{j,i}\}\rangle$, and*
- *an integer $n < N$.*

Among all samples $I \subseteq \{1, \ldots, N\}$ of size n, we must find the sample I for which the distance $\rho(t^{(2)}(I), t^{(2)})$ between the corresponding (E, V)-statistical tuples is the smallest possible.

Definition 10. *Let ρ be a distance function, and let $d \geq 1$ be an integer. By a* $2d$-th order sample selection problem *corresponding to ρ, we mean the following problem. We are given:*

- *a population $p = \langle N, k, \{x_{j,i}\}\rangle$, and*
- *an integer $n < N$.*

Among all samples $I \subseteq \{1, \ldots, N\}$ of size n, we must find the sample I for which the distance $\rho(t^{(2d)}(I), t^{(2d)})$ between the corresponding $(2d)$-th order statistical tuples is the smallest possible.

Proposition 1. *For every distance function ρ, the corresponding E-sample selection problem is NP-hard.*

Proposition 2. *For every distance function ρ, the corresponding (E, V)-sample selection problem is NP-hard.*

Proposition 3. *For every distance function ρ and for every integer $d \geq 1$, the corresponding $(2d)$-th order sample selection problem is NP-hard.*

What is NP-hardness: a brief informal reminder. In order to prove these results, let us recall what NP-hardness means. Informally, a problem $\mathcal{P}_0$ is called *NP-hard* if it is at least as hard as all other problems from the class NP (a natural class of problems).

To be more precise, a problem $\mathcal{P}_0$ is NP-hard if every problem $\mathcal{P}$ from the class NP can be reduced to this problem $\mathcal{P}_0$. A *reduction* means that to every instance p of the problem $\mathcal{P}$, we must be able to assign (in a feasible, i.e., polynomial-time way) an instance p_0 of our problem $\mathcal{P}_0$ in such a way that the solution to the new instance p_0 will lead to the solution of the original instance p. For precise definitions, see, e.g., [19].

How NP-hardness is usually proved. The original proof of NP-hardness of certain problems $\mathcal{P}_0$ is rather complex, because it is based on explicitly proving that *every* problem from the class NP can be reduced to the problem $\mathcal{P}_0$. However, once we have proven NP-hardness of a problem $\mathcal{P}_0$, the proof of NP-hardness of other problems $\mathcal{P}_1$ is much easier.

Indeed, from the above description of a reduction, one can easily see that reduction is a transitive relation: if a problem $\mathcal{P}$ can be reduced to a problem $\mathcal{P}_0$, and the problem $\mathcal{P}_0$ can be reduced to a problem $\mathcal{P}_1$, then, by combining these two reductions, we can prove that $\mathcal{P}$ can be reduced to $\mathcal{P}_1$.

Thus, to prove that a new problem $\mathcal{P}_1$ is NP-hard, it is sufficient to prove that one of the known NP-hard problems $\mathcal{P}_0$ can be reduced to this problem $\mathcal{P}_1$. Indeed, since $\mathcal{P}_0$ is NP-hard, every other problem $\mathcal{P}$ from the class NP can be reduced to this problem $\mathcal{P}_0$. Since $\mathcal{P}_0$ can be reduced to $\mathcal{P}_1$, we can now conclude, by transitivity, that every problem $\mathcal{P}$ from the class NP can be reduced to this problem $\mathcal{P}_1$ – i.e., that the problem $\mathcal{P}_1$ is indeed NP-hard.

Comment. As a consequence of the definition of NP-hardness, we can conclude that if a problem $\mathcal{P}_0$ is NP-hard, then every more general problem $\mathcal{P}_1$ is also NP-hard.

Indeed, the fact that $\mathcal{P}_0$ is NP-hard means that every instance p of every problem $\mathcal{P}$ can be reduced to some instance p_0 of the problem $\mathcal{P}_0$. Since the problem $\mathcal{P}_1$ is more general than the problem $\mathcal{P}_0$, every instance p_0 of the problem $\mathcal{P}_0$ is also an instance of the more general problem $\mathcal{P}_1$.

Thus, every instance p of every problem $\mathcal{P}$ can be reduced to some instance p_0 of the problem $\mathcal{P}_1$ – i.e., that the more general problem $\mathcal{P}_1$ is indeed NP-hard.

Main idea of the proof: reduction to subset sum, a known NP-hard problem. We prove NP-hardness of our problem by reducing a known NP-hard problem to it: namely, a *subset sum* problem, in which we are given m positive integers $s_1, \ldots, s_m$, and we must find the signs $\varepsilon_i \in \{-1, 1\}$ for which

$$\sum_{i=1}^{m} \varepsilon_i \cdot s_i = 0;$$

see, e.g., [19].

A reduction means that to every instance $s_1, \ldots, s_m$ of the subset sum problem, we must assign (in a feasible, i.e., polynomial-time way) an instance of our problem in such a way that the solution to the new instance will lead to the solution of the original instance.

Reduction: explicit description. Let us describe this reduction: we take $N = 2n$, $k = 2$, $n = m$, and we select the values $x_{j,i}$ as follows:

- $x_{1,i} = s_i$ and $x_{1,m+i} = -s_i$ for all $i = 1, \ldots, m$;
- $x_{2,i} = x_{2,m+i} = 2^i$ for all $i = 1, \ldots, m$.

We will show that for this new problem, the most representative sample I has $\rho(t(I), t) = 0$ if and only if the original instance of the subset sum problem has a solution.

General analysis. Indeed, by definition of a distance function, the equality $\rho(t(I), t) = 0$ is equivalent to $t(I) = t$, i.e., to the requirement that for the sample I, means (and variances) within the sample are exactly the same as for the entire population.

Consequences for the second component. Let us start by analyzing the consequences of this requirement for the mean of the second component. For the entire population of size $N = 2m$, for each i from 1 to m, we have two elements, i-th and $(m + i)$-th, with the value $x_{2,i} = x_{2,m+i} = 2^i$. Thus, for the population as a whole, this mean is equal to

$$E_2 = \frac{2 + 2^2 + \ldots + 2^m}{m}.$$

For the selected subset I of size m, this mean should be exactly the same: $E_2(I) = E_2$. Thus, we must have

$$E_2(I) = \frac{2 + 2^2 + \ldots + 2^m}{m}.$$

By definition,

$$E_2(I) = \frac{1}{m} \cdot \sum_{i \in I} x_{2,i}.$$

Thus, we conclude that

$$S_2(I) \stackrel{\text{def}}{=} \sum_{i \in I} x_{2,i} = 2 + 2^2 + \ldots + 2^m.$$

What can we now conclude about the set I?

First of all, we can notice that in the sum $2+2^2+\ldots+2^m$, all the terms are divisible by 4 except for the first term 2. Thus, the sum itself is not divisible by 4.

In our population, we have exactly two elements, element 1 and element $m + 1$, for which $x_{2,1} = x_{2,m+1} = 2$. For every other element, we have $x_{2,i} = x_{2,m+i} = 2^i$ for $i \geq 2$ and therefore, the corresponding value is divisible by 4.

In regards to a selection I, there are exactly three possibilities:

- the set I contains none of the two elements 1 and $m+1$;
- the set I contains both elements 1 and $m+1$; and
- the set I contains exactly one of the two elements 1 and $m+1$.

In the first two cases, the contribution of these two elements to the sum $S_2(I)$ is divisible by 4 (it is 0 or 4). Since all other elements in the sum $S_2(I)$ are divisible by 4, we would thus conclude that the sum itself is divisible by 4 – which contradicts to our conclusion that this sum is equal to $2+2^2+\ldots+2^m$ and is, therefore, not divisible by 4.

This contradiction shows that the set I must contain exactly one of the two elements 1 and $m+1$. Let us denote this element by k_1. For this element, $x_{2,k_1}=2$. Subtracting x_{2,k_1} and 2 from the two sides of the equality

$$S_2(I)=\sum_{i\in I} x_{2,i}=2+2^2+\ldots+2^m,$$

we conclude that

$$S_2(I-\{k_1\})=\sum_{i\in I-\{k_1\}} x_{2,i}=2^2+2^3+\ldots+2^m.$$

In the new sum $2^2+2^3+\ldots+2^m$, all the terms are divisible by $2^3=8$ except for the first term 2^2. Thus, the sum itself is not divisible by 8.

In our remaining population $\{2,\ldots,m,m+2,\ldots,2m\}$, we have exactly two elements, element 2 and element $m+2$, for which $x_{2,2}=x_{2,m+2}=2^2$. For every other element, we have $x_{2,i}=x_{2,m+i}=2^i$ for $i\geq 3$ and therefore, the corresponding value is divisible by 3.

In regards to a selection I, there are exactly three possibilities:

- the set I contains none of the two elements 2 and $m+2$;
- the set I contains both elements 2 and $m+2$; and
- the set I contains exactly one of the two elements 2 and $m+2$.

In the first two cases, the contribution of these two elements to the sum $S_2(I-\{k_1\})$ is divisible by 8 (it is 0 or 8). Since all other elements in the sum $S_2(I-\{k_1\})$ are divisible by 8, we would thus conclude that the sum itself is divisible by 8 – which contradicts to our conclusion that this sum is equal to $2^2+2^3+\ldots+2^m$ and is, therefore, not divisible by 8.

This contradiction shows that the set I must contain exactly one of the two elements 2 and $m+2$. Let us denote this element by k_2. For this element, $x_{2,k_2}=2^2$. Subtracting x_{2,k_2} and 2^2 from the two sides of the equality

$$S_2(I-\{k_1\})=\sum_{i\in I-\{k_1\}} x_{2,i}=2^2+2^3+\ldots+2^m,$$

we conclude that

$$S_2(I-\{k_1,k_2\})=\sum_{i\in I-\{k_1,k_2\}} x_{2,i}=2^3+2^4+\ldots+2^m.$$

Now, we can similarly conclude that the set I contains exactly one element from the pair $\{3, m+3\}$, and in general, for every i from 1 to m, we can conclude that the selection set I contains exactly one element k_i from the pair $\{i, m+i\}$.

Consequences for the first component. Let us now analyze the consequences of this requirement for the mean of the first component. For the entire population of size $N = 2m$, for each i from 1 to m, we have two elements, i-th and $(m+i)$-th, with the opposite values $x_{1,i} = s_i$ and $x_{2,m+i} = -s_i$. Thus, for the population as a whole, this mean is equal to $E_1 = 0$.

For each i from 1 to m, the selection set contains exactly one element of these two: $k_i = i$ and $k_i = m+i$. Thus, $E_1(I) = 0$ means that the corresponding sum is equal to 0: $\sum_{i=1}^{m} x_{1,k_i} = 0$. Here, $x_{1,k_i} = \varepsilon_i \cdot s_i$, where:

- $\varepsilon_i = 1$ if $k_i = i$, and
- $\varepsilon_i = -1$ if $k_i = m+i$.

Thus, we conclude that $\sum_{i=1}^{m} \varepsilon_i \cdot s_i = 0$ for some $\varepsilon_i \in \{-1, 1\}$, i.e., that the original instance of the subset problem has a solution.

Equivalence. Vice versa, if the original instance of the subset problem has a solution, i.e., if $\sum_{i=1}^{m} \varepsilon_i \cdot s_i = 0$ for some $\varepsilon_i \in \{-1, 1\}$, then we can select $I = \{k_1 \ldots, k_m\}$, where:

- $k_i = i$ when $\varepsilon_i = 1$, and
- $k_i = m+i$ when $\varepsilon_i = -1$.

One can easily check that in this case, we have $E_1(I) = E_1$, $E_2(I) = E_2$, $V_1(I) = V_1$, $V_2(I) = V_2$, and, in general, $M_1^{(2d)}(I) = M_1^{(2d)}$ and $M_2^{(2d)}(I) = M_2^{(2d)}$.

Conclusion. The reduction is proven, so the problem of finding the most representative sample is indeed NP-hard.

Discussion. In the definitions of sample selection problem $\mathcal{P}_1$ (Definitions 8–10), the objective is to find the sample I of given size n (which is smaller than N, the size of the population) such that the distance $\rho(t(I), t)$ is the smallest possible.

In the above text, we have proved, in effect, that the selection of a sample I of a given size n ($< N$), such that the distance $\rho(t(I), t) = 0$, is NP-hard.

The distance is always non-negative. Thus, when the smallest possible distance is 0, finding the sample I for which the distance $\rho(t(I), t)$ is the smallest possible is equivalent to finding the sample for which this distance is zero. In general, the smallest possible distance does not necessarily equal to 0. Thus, the sample selection problem $\mathcal{P}_1$ is more general that the auxiliary "zero-distance" problem $\mathcal{P}_0$ for which we have proven NP-hardness.

We have already mentioned earlier that if a problem $\mathcal{P}_0$ is NP-hard, then a more general problem $\mathcal{P}_1$ is NP-hard as well. Thus, we have indeed proved that the (more general) sample selection problem is NP-hard.

Towards auxiliary results. In our proofs, we considered the case when the desired sample contains half of the original population. In practice, however, samples form a much smaller portion of the population. A natural question is: what if we fix a large even number $2P \gg 2$, and look for samples which constitute the $(2P)$-th part of the original population? It turns out that the resulting problem of selecting the most representative sample is still NP-hard.

Definition 11. *Let ρ be a distance function, and let $2P$ be a positive even integer. By a* problem of selecting an E-sample of relative size $\dfrac{1}{2P}$*, we mean the following problem:*

- *We are given a population $p = \langle N, k, \{x_{j,i}\}\rangle$.*
- *Among all samples $I \subseteq \{1, \ldots, N\}$ of size $n = \dfrac{n}{2P}$, we must find the sample I for which the distance $\rho(t^{(1)}(I), t^{(1)})$ between the corresponding E-statistical tuples is the smallest possible.*

Definition 12. *Let ρ be a distance function, and let $2P$ be a positive even integer. By a* problem of selecting an (E, V)-sample of relative size $\dfrac{1}{2P}$*, we mean the following problem:*

- *We are given a population $p = \langle N, k, \{x_{j,i}\}\rangle$.*
- *Among all samples $I \subseteq \{1, \ldots, N\}$ of size $n = \dfrac{n}{2P}$, we must find the sample I for which the distance $\rho(t^{(2)}(I), t^{(2)})$ between the corresponding (E, V)-statistical tuples is the smallest possible.*

Definition 13. *Let ρ be a distance function, let $d \geq 1$ be an integer, and let $2P$ be a positive even integer. By a* problem of selecting an $(2d)$-th order sample of relative size $\dfrac{1}{2P}$*, we mean the following problem:*

- *We are given a population $p = \langle N, k, \{x_{j,i}\}\rangle$.*
- *Among all samples $I \subseteq \{1, \ldots, N\}$ of size $n = \dfrac{n}{2P}$, we must find the sample I for which the distance $\rho(t^{(2d)}(I), t^{(2d)})$ between the corresponding statistical tuples of order $2d$ is the smallest possible.*

Proposition 4. *For every distance function ρ and for every even integer $2P$, the corresponding problem of selecting an E-sample of relative size $\dfrac{1}{2P}$ is NP-hard.*

Proposition 5. *For every distance function ρ and for every even integer $2P$, the corresponding problem of selecting an (E, V)-sample of relative size $\dfrac{1}{2P}$ is NP-hard.*

Proposition 6. *For every distance function ρ, for every integer $d \geq 1$, and for every even integer $2P$, the corresponding problem of selecting a $(2d)$-th order sample of relative size $\dfrac{1}{2P}$ is NP-hard.*

Proof of Propositions 4–6. The proof is similar to the proofs of Propositions 1–3.

The main difference is that for each i from 1 to m, we now have not two but $2P$ different objects

$$i, m+i, 2m+i, \ldots, k \cdot m + i, \ldots, (2P-1) \cdot m + i$$

with the same value

$$x_{2,i} = x_{2,m+i} = \ldots = x_{2,k \cdot m+i} = \ldots = x_{2,(2P-1) \cdot m+i} = (2P)^i.$$

(And this common value is also different.)

Among these $2P$ objects with the same value of the second characteristic $x_{2,\cdot}$, for the first half, we have $x_{1,\cdot} = s_i$ and for the second half, we have $x_{1,\cdot} = -s_i$, i.e.:

$$x_{1,i} = x_{1,m+i} = \ldots = m_{1,(P-1) \cdot m+i} = s_i;$$

$$x_{1,P \cdot m+i} = x_{1,(P+1) \cdot m+i} = \ldots = m_{1,(2P-1) \cdot m+i} = -s_i.$$

By using divisibility by $(2P)^2$ (instead of divisibility by 2^2), we conclude that the best fitting sample is the one which has exactly one element of each group. Thus, from $E_1(I) = E_1$, we similarly conclude that the original instance of the subset problem has a solution – and hence that the new problems are indeed NP-hard.

7 Symmetry: Another Fundamental Reason for Continuity ("Fuzziness")

Case study: benzene. To explain why symmetry leads to continuity, let us start with a chemical example. In the traditional chemistry, a molecule is composed from atoms that exchange electrons with each other. If an atom borrows one electron from another atom, we say that they have a connection of valence 1, if two electrons, there is a connection of valence 2, etc.

From the analysis of benzene, it has been clear that it consists of 6 carbon and six hydrogen atoms, i.e., that its chemical formula is C_6H_6. However, for a long time, it was not clear how exactly they are connected to each other. The solution came in the 19th century to a chemist August Kekule in a dream. He dreamed of six monkeys that form a circle in which each monkey holds to the previous monkey's tail. According to this solution, the six C atoms form a circle. To each of these atoms, a H atom is attached. Each C atom has a 1

valence connection to H, 1 valence connection to one of its neighbors, and 2 to another neighbor.

The resulting chemical structure is still routinely described in chemical textbooks – because a benzene loop is a basis of organic chemistry and life. However, now we understand that this formula is not fully adequate. Indeed, according to this formula, the connections between C atoms are of two different types: of valence 1 and of valence 2. In reality, the benzene molecule is completely symmetric, there is no difference between the strengths of different connections.

It is not possible to have a symmetric configuration is we require that valencies are integers. To equally split the remaining valence of 3 (1 is taken for H) between the two neighbors, we need a valence of 3/2. This is not possible in classical chemistry – but this is possible, in some sense, in quantum chemistry where, as we have mentioned, we have a continuum of intermediate states; see, e.g., [2].

Fuzzy logic itself is such an example. Fuzzy logic itself can be viewed as an example where symmetries leads to values intermediate between the original discrete values.

Indeed, in traditional logic, we have two possible truth values: 1 ("true") and 0 ("false"). How can we use this logic to describe the absence of knowledge? If we do not know whether a given statement A is true or not, this means that we have the exact same degree of belief in the statement A as we have in its negation $\neg A$. In the traditional logic, none of the two truth values are symmetric (invariant) under such transformation $A \to \neg A$. Thus, to adequately describe this situation, we need to also consider additional (intermediate) truth values.

And indeed, in fuzzy logic with the set of truth values $[0, 1]$ and the negation operation $f_\neg(x) = 1 - x$, there is a value which is invariant under the operation $A \to \neg A$: the value 0.5.

8 Case Study: Territory Division

Formulation of the problem. In many conflict situations, several participants want to divide a territory between themselves. It may be farmer's children dividing his farm, it may be countries dividing a disputed territory.

Traditional (non-fuzzy) formalization of the problem. Let us follow [15] and describe a traditional (non-fuzzy) formalization of this problem. Let us denote the disputed territory (i.e., to be more precise, the set of all the points in this territory) by T. Our objective is to divide this territory between n participants, i.e., to select a division of the set T into the sets $T_1, T_2, \ldots, T_n$ for which $T_i \cap T_j = \emptyset$ for $i \neq j$ and

$$T_1 \cup T_2 \cup \ldots \cup T_n = T.$$

It is reasonable to assume that the utility u_i of the i-th participant in acquiring the territory T_i is linear in T_i, i.e., has the form

$$u_i(T_i) = \int_{T_i} U_i(x)\,dx$$

for some appropriate function $U_i(x)$. As we mentioned in [15], it is reasonable to use Nash's criterion to select the optimal division, i.e., to select the division for which the product

$$u \stackrel{\text{def}}{=} u_1(T_1) \cdot u_2(T_2) \cdot \ldots \cdot u_n(T_n)$$

attains the largest possible value. According to [15], in the optimal solution, for every participants i, there is a weight c_i such that each point x is assigned to the participant with the largest weighted utility $c_i \cdot U_i(x)$.

In particular, for two participants, there is a threshold c such that all the points x for which $U_1(x)/U_2(x) > c$ go to the first participant, and all the points x for which $U_1(x)/U_2(x) < c$ go to the second participant.

Possibility of a "fuzzy" solution. From the commonsense viewpoint, why do we have to necessarily divide all the disputed territory? Why cannot we control some parts of it together? In other words, instead of dividing the set T into subsets T_i, why cannot we assign, to every point $x \in T$ and to every i, the degree $d_i(x)$ to which the i-th participant will control the neighborhood of this point – in such a way that for every point x,

$$d_1(x) + \ldots + d_n(x) = 1.$$

In other words, instead of a crisp partition we have a fuzzy partition.

In this setting, the utility u_i of the i-th participant has the form

$$u_i(d_i) = \int U_i(x) \cdot d_i(x)\,dx,$$

and our objective is to find a fuzzy partition for which the product

$$u \stackrel{\text{def}}{=} u_1(d_1) \cdot u_2(d_2) \cdot \ldots \cdot u_n(d_n)$$

attains the largest possible value.

Observation: the above "fuzzy" problem always has a crisp optimal solution. The derivation from [15] was based on the idea that if we attain a maximum, then a small change of assignment in the vicinity of each point will only decrease (or not change) the desired product. For the fuzzy problem, a similar argument shows that there are weights c_i such that in the optimal solution, every point x for which the weighted utility each point x is assigned to the participant with the largest weighted utility $c_i \cdot U_i(x)$ of the i-th participant is larger than the weighted utility of all other participants is assigned to this i-th participant.

The only points about which we cannot make a definite assignment are the ones in which two or more participants have exactly the same weighted utility. How we divide these points between these participants does not matter – as long as the overall degree of all the points assigned to each of these participants remains the same. In particular, this means that it is always possible to have a crisp division with the optimal value of the desired product.

So, we arrive at a somewhat paradoxical situation: even when we allow "fuzzy" divisions, the corresponding optimization problem always have a crisp solution. So, at first glance, it may seem that fuzzy solutions are not needed at all.

As we will see, the situation changes if we consider symmetry.

Symmetry leads to fuzziness. For the territory division problem, a symmetry means a transformation $f : T \to T$ that preserves the area of each (crisp) subset and that preserves the utility of each subarea to each participant. Preserving area means that f has to be a measure-preserving transformation. Preserving utility means that we must have $U_i(x) = U_i(f(x))$ for all x.

It is reasonable to require that if the original situation allows a symmetry, then the desired division should be invariant with respect to this symmetry. Let us show that this requirement leads to a fuzzy solution.

Indeed, let us consider the simplest situation in which we have only two participants, and both assign equal value to all the points $U_1(x) = U_2(x) = 1$. In this case, the utility of each set T_i is simply equal to its area A_i, so the optimization problem takes the form

$$A_1 \cdot A_2 \to \max .$$

Since the sum $A_1 + A_2$ is equal to the area A of the original territory T, this problem takes the form

$$A_1 \cdot (A - A_1) \to \max .$$

One can easily check that the optimal crisp solution means that $A_1 = A/2$, i.e., that we divide the area T into two equal halves.

This solution is optimal but it is not symmetric. Indeed, in this case, symmetries are simply area-preserving transformations. Symmetry of the division means that $f(T_1) = T_1$ for all such transformations f. However, for every two points $x, y \in T$, we can have an area-preserving transformation f that maps x into y: $f(x) = y$. In particular, we can have sauch a transformation for $x \in T_1$ and $y \in T_2$, in which case $f(T_1) \neq T_1$. Thus, a crisp symmetric solution is impossible.

In contrast, a fuzzy symmetric solution is quite possible – and uniquely determined: we simply assign to each point x equal degrees $d_1(x) = d_2(x) = 1/2$. Then, $f(d_1) = d_1$ and $f(d_2) = d_2$ for all area-preserving transformations f.

In general, we always have an optimal symmetric solution: in this solution, equally desired points – for which $c_i \cdot U_i(x) = c_j \cdot U_j(x)$ – are all assigned a joint control with the same degree of ownership depending only on i and j.

9 Conclusion

In this paper, we have proven that from the natural assumption that the world is cognizable, we can conclude that intermediate degrees are needed to describe real-world processes. This conclusion provides an additional explanation for the success of fuzzy techniques (and other techniques which use intermediate degrees) – success which often goes beyond situations in which the intermediate degrees are needed to describe the experts' uncertainty.

Acknowledgment

This work was supported in part by NSF grants HRD-0734825, EAR-0225670, and EIA-0080940, by Texas Department of Transportation grant No. 0-5453, by the Japan Advanced Institute of Science and Technology (JAIST) International Joint Research Grant 2006-08, and by the Max Planck Institut für Mathematik.

References

1. S. Basu, R. Pollack, and M.-F. Roy, *Algorithms in Real Algebraic Geometry*, Springer-Verlag, Berlin, 2006.
2. R. P. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, 2005.
3. T. Franzen, *Gödel's Theorem: An Incomplete Guide to its Use and Abuse*, A. K. Peters, 2005.
4. J. E. Gamez, F. Modave, and O. Kosheleva, "Selecting the Most Representative Sample is NP-Hard: Need for Expert (Fuzzy) Knowledge", *Proceedings of the IEEE World Congress on Computational Intelligence WCCI'2008*, Hong Kong, China, June 1–6, 2008, pp. 1069–1074.
5. M. Gell-Mann, *The Quark and the Jaguar*, Owl Books, 1995.
6. L. K. Grover, "A fast quantum mechanical algorithm for database search", *Proceedings of the 28th Annual ACM Symposium on the Theory of Computing*, May 1996, pp. 212-ff.
7. L. K. Grover, "From Schrödinger's equation to quantum search algorithm", *American Journal of Physics*, 2001, Vol. 69, No. 7, pp. 769–777, 2001.
8. G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications.* Prentice Hall, Upper Saddle River, New Jersey, 1995.
9. C. C. Kuo, F. Glover, and K. S. Dhir, "Analyzing and modeling the maximum diversity problem by zero-one programming", *Decision Sciences*, vol. 24, no. 6, pp. 1171–1185, 1993.
10. H. Lohr, *Sampling: Design and Analysis*, Duxbury Press, 1999.
11. B. Mandelbrot, *Fractals: Form, Chance and Dimension*, W. H. Freeman and Co, 1977.
12. B. Mandelbrot, *The Fractal Geometry of Nature.* W. H. Freeman & Co, 1982.

13. B. Mandelbrot and and R. L. Hudson, *The (Mis)Behavior of Markets: A Fractal View of Risk, Ruin, and Reward*, Basic Books, 2004.
14. B. Mishra, "Computational real algebraic geometry", in: *Handbook on Discreet and Computational Geometry*, CRC Press, Boca Raton, Florida, 1997.
15. H. T. Nguyen and V. Kreinovich, "How to Divide a Territory? A New Simple Differential Formalism for Optimization of Set Functions", *International Journal of Intelligent Systems*, 1999, Vol. 14, No. 3, pp. 223–251.
16. H. T. Nguyen and V. Kreinovich, "Everything Is a Matter of Degree: A New Theoretical Justification of Zadeh's Principle", *Proceedings of the 27th International Conference of the North American Fuzzy Information Processing Society NAFIPS'2008*, New York, New York, May 19–22, 2008.
17. H. T. Nguyen and E. A. Walker, *A first course in fuzzy logic*, CRC Press, Boca Raton, Florida, 2005.
18. M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, Cambridge, 2000.
19. C. H. Papadimitriou, *Computational Complexity*, Addison Wesley, San Diego, 1994.
20. B. Povh, *Particles and Nuclei: An Introduction to the Physical Concepts*, Springer-Verlag, 1995.
21. P. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer", *Proceedings of the 35th Annual Symposium on Foundations of Computer Science*, Santa Fe, NM, Nov. 20–22, 1994.
22. P. Shor, "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer", *SIAM J. Sci. Statist. Comput.*, 1997, Vol. 26, pp. 1484-ff.
23. A. Tarski, *A Decision Method for Elementary Algebra and Geometry*, 2nd ed., Berkeley and Los Angeles, 1951, 63 pp.