8-1998

Cooperative Learning is Better: Explanation Using Dynamical Systems, Fuzzy Logic, and Geometric Symmetries

Vladik Kreinovich
The University of Texas at El Paso, vladik@utep.edu

Edye Johnson-Holubec

Leonid K. Reznik

Misha Kosheleva

Comments:
Technical Report: UTEP-CS-98-24
In: Nguyen Hoang Phuong and Ario Ohsato (eds.), *Proceedings of the Vietnam-Japan Bilateral Symposium on Fuzzy Systems and Applications VJFUZZY'98*, HaLong Bay, Vietnam, 30th September-2nd October, 1998, pp. 154-160.

Cooperative Learning is Better: Explanation Using Dynamical Systems, Fuzzy Logic, and Geometric Symmetries

Vladik Kreinovich[1], Edye Johnson-Holubec[2], Leonid K. Reznik[3], and Misha Koshelev[4]

[1]Department of Computer Science, University of Texas at El Paso
El Paso, TX 79968, USA, email vladik@cs.utep.edu

[2]Cooperative Learning Center, P.O. Box 552
Taylor, TX 76574, USA, email edyeh@totalaccess.net

[3]Department of Electrical and Electronic Engineering
Footscray Campus, Victoria University of Technology
MMC Melbourne VIC 8001, Australia, email LeonReznik@vut.edu.au

[4]1003 Robinson, El Paso, TX 79902, USA, email mkosh@cs.utep.edu

1 Introduction

Cooperative learning is efficient. It is well-known that when students help each other by studying in (well-organized and well-monitored) small groups, then they learn faster and more efficiently than when they study in a (more traditional) individual manner. Cooperative learning is widely spread in Japan, and it has been re-invented and actively used in the USA (see, e.g., [4, 5] and references therein).

In spite of empirical success of cooperative learning, there are few theoretical explanations. There is a lot of literature on cooperative learning. The corresponding statistics shows that cooperative learning (if properly administered) is indeed better, and provides the empirical recommendations on how to choose the optimal parameters of cooperative learning. For example:

- the optimal group size is around 2 or 3, and
- the optimal percentage of class time spent on cooperative mode of learning is around 60%.

There are, however, no theoretical models which would explain this empirical data.

Our results. We show that the combination of fuzzy and non-fuzzy techniques can design such models. Namely, we do the following:

- first, we use the dynamical system model to provide a simple explanation of why cooperative learning is more efficient;
- second, we use the ideas from a fuzzy explanation of golden proportion [8] to explain why allocating 60% of class time to cooperative learning is the optimal strategy;
- finally, we use the general geometric symmetry approach developed in [11] (approach that explains the most frequently used fuzzy and neural heuristics) to find the optimal student placement and thus, to show that groups of 2 or 3 are indeed optimal.

2 Dynamical system model explains why cooperative learning is more efficient

At first glance, it may seem that cooperative learning, in which students learn from each other as they study, is not necessarily a good idea: it is better to learn from a knowledgeable professor than from a fellow student who is approximately at the same level of knowledge (and who definitely knows much less than a professor). However, experiments show that when implemented appropriately, cooperative learning does help. Why?

In this section, we will show that even the simplest dynamical system models explain why cooperative learning is more efficient. We will start with an oversimplified linear model, and then show that in a more realistic model, cooperative learning is indeed more efficient.

Learning as a dynamical system. Before we start the explanations, let us describe why dynamical system models are indeed a good description of the learning process.

A dynamical system is a general mathematical description of systems which change in time. Learning, by definition, means changing the state of student's knowledge, so it is very natural to use dynamical systems to describe learning.

Learning means changing the original state s_0 of the learner's mind into a desired "learned" state $s = f(s_0)$. Learning consists of several "elementary steps" such a memorizing, testing, listening, etc. There is only so much that can be achieved via one "learning step": we can learn a simple fact or a simple algorithm in one step, but most frequently, we need a *sequence* of steps to learn.

First, we describe the class F of functions which can be achieved in a single learning step. Functions f belonging to this class can be learned in a single step. If the desired function f does not belong to the class F, but we can try to learn this function in *two* steps. If we choose a first step which corresponds to a function g, and a second step which corresponds to a function h, then:

- the original state $s(0) = s_0$ is transformed first into state $s(1) = g(s_0)$, and
- then into a state $s(2) = h(s(1)) = h(g(s_0))$ which corresponds to the function $f(s) = g(h(s))$.

Thus, in two steps, we can implement all functions which can be represented as compositions of two one-step functions. Similarly, in three steps, we can implement all functions which can be represented as compositions of three one-step functions, etc.

Cooperative learning as a dynamical system. How can we represent cooperative learning in this framework? Cooperative learning means that the next step of each learner depends not only on the previous state of *this* particular learner, but also on the previous state of all *other* learners. Thus, if we have p cooperating learners, and we want to describe how their states $s_1, \ldots, s_p$ change in time, from the states $s_i(t)$ at the moment of time t to the states $s_i(t+1)$ at the next moment of time $t+1$, we must describe p functions of p variables: $s_i(t+1) = f_i(s_1(t), \ldots, s_p(t))$. So, to analyze the possibilities of cooperative learning, we must describe the class F_p of functions of p state variables (as opposed to functions of one state variable for non-cooperative learning).

We will show that this cooperation can indeed improve the learning ability, even if we want all the learners to learn the same material (or the same skills), i.e., if we want them to learn the same transformation $s_0 \to s$.

Choosing the simplest model. To determine the state of a dynamical system, we must undertake some measurements. As a result of these measurements, we get the values of the measured quantities which characterize the current state. In other words, from the mathematical viewpoint, a state s can be represented by a finite sequence $q_1, \ldots, q_m$ of real numbers (values of the measured quantities). Therefore, in the following text, we will identify a state s with such a sequence $q_1, \ldots, q_m$. Correspondingly, functions from states to states are transformations from R^m to R^m.

Which transformations should we choose? The simplest possible transformations are *linear* transformations, so it may seem natural to choose, as F, the class of all linear transformations from R^m to R^m. However, this model is oversimplified: since the composition of linear transformations is always linear, in this model, whatever we can learn in two, three, etc. learning steps, we can also learn in a single learning step as well.

Thus, if we want a more realistic model, in which in two learning steps, we can potentially learn more than in a single learning step, then we must choose a more complicated set of one-step transformations F. After linear functions, the next natural class is the class of all *bilinear* transformations, i.e., transformations which are linear in each of its variables. For bilinear transformations, in two steps, we can already get more than in a single step: e.g., for $m = 2$, the transformation $f : R^2 \to R^2$ defined as $f_1(q_1, q_2) = f_2(q_1, q_2) = q_1^2$ is not itself bilinear and therefore, cannot be learned in a single step, but it can be learned in two steps:

- first, we apply a bilinear transformation
$$g : (q_1, q_2) \to (q_1, q_1);$$
- then, we apply another bilinear transformation
$$h : (q_1, q_2) \to (q_1 \cdot q_2, q_1 \cdot q_2).$$

After these two transformations, we get

$$q_1(t+1) = q_2(t+1) = q_1(t)$$

and

$$q_2(t+1) = q_2(t+1) = q_1(t+1) \cdot q_2(t+1) = q_1^2(t),$$

i.e., the desired transformation f.

Therefore, as the simplest model of learning, we will take the class F of all bilinear transformations.

Similarly, to describe cooperative learning, we will consider, for every p, the set of all p-tuples of quadratic functions of $R^m \times \ldots \times R^m$ (p times) $\to R^m$.

Cooperative learning is indeed more efficient: a proof. Of course, whatever we can learn individually, we can always learn cooperatively within the same time period if we simply do not cooperate.

Let us show that, within the above model, for each m, cooperative learning is indeed more efficient: namely, we will show that there exist tasks (i.e., functions) which can be learned in two learning steps if we use cooperative learning, but which require more learning steps if we do not use cooperation.

It is sufficient to show this result for $p = 2$, because if it is true for $p = 2$ (i.e., for groups of two), then to prove that cooperative learning is efficient for $p > 2$, we can divide all p learners into groups of two, and let each pair learn more efficiently (if p is odd, we can add the remaining learner to one of the pairs). In view of this comment, in the following text, we will assume that $p = 2$.

Without cooperation, since each of the transformations is bilinear (hence quadratic), in two steps, we can only get quartic transformations, i.e., transformations which are expressed by polynomials of $\le$ 4-th order.

With cooperation, we can get, in two steps, the following higher order transformation: $(q_1, q_2, \ldots) \to (q_1^4 \cdot q_2^4, 0, \ldots, 0)$. Indeed, to get this 8-th order transformations, we can, on the first step, use, for both learners with initial states $s_1(0) = s_2(0) = (q_1, q_2, \ldots)$, a non-cooperative bilinear transformation

$$g_1(q_1, q_2, \ldots) = \ldots = g_m(q_1, q_2, \ldots) = q_1 \cdot q_2$$

(which does not depend on the state of the other learner at all), and then, on the second step, use, for each learner, a bilinear transformation $s = h_1(s_1, s_2) = h_2(s_1, s_2)$ which is defined as follows: for $s_1 = (q_{11}, q_{12}, \ldots, q_{1m})$ and $s_2 = (q_{21}, q_{22}, \ldots, q_{2m})$, we define the resulting state $s_r = (q_{r1}, \ldots, q_{rm})$ as $q_{1r} = q_{11} \cdot q_{12} \cdot q_{21} \cdot q_{22}$ and $q_{ir} = 0$ for all $i \ge 2$. In this case, if we start with the state $s_i(t) = (q_1, q_2, \ldots)$ of both learners, after the first learning step, both learners get into the state $s_i(t+1) = (q_1 \cdot q_2, q_1 \cdot q_2, \ldots, q_1 \cdot q_2)$, and after the second learning step, they both get into the desired state $s_i(t+2) = (q_1^4 \cdot q_2^4, 0, \ldots, 0)$. The statement is proven.

Comment. Our proof does not depend on whether we have human or robotic learners. Thus, it not only shows that cooperative learning is better for *human* learners, it also shows that it is more efficient for automated *intelligent agents* as well.

3 Fuzzy logic explains why allocating 60% of class time to cooperative learning is the optimal strategy

Cooperation is advantageous, but some learning is best done individually. What portion x of class time should be allocated to cooperative learning and what portion to individual learning? If we start with 0 portion $x = 0$ (no collaboration at all), and gradually increase it (i.e., add some cooperation), we will make the learning more efficient. However, after a certain value x, we will get a decrease in efficiency. The optimal portion allocated for cooperative learning can be described by the following condition: further increase in this degree leads to an opposite effect. We can rewrite this condition as follows:

$$\text{"very" } x = \text{ "not" } x.$$

To formalize "very" and "not", it is natural to use fuzzy logic [6, 12], where:

- "very" x is typically interpreted as x^2, and
- "not" x is usually interpreted as $1 - x$.

> *Historical comment.* The interpretation of "very" as x^2 was originally proposed by L.A. Zadeh in his pioneer paper [15]; the experimental results of [7] turned out to be consistent with this interpretation of "very".

If we use these interpretations in the above formula, we get the equation $x^2 = 1 - x$, whose only solution of the interval $[0, 1]$ is the *golden proportion* number $(\sqrt{5}-1)/2 = 0.618\ldots \approx 0.6$. This portion is consistent with the experimental results according to which the optimal portion is $\approx 60\%$. Thus, fuzzy logic indeed explains this experimental result.

4 General geometric symmetry approach explains why groups of 2 or 3 students are optimal

In order to answer the question of what is the optimal class size, let us consider a related important question: what is the optimal student placement? This question is very important because the whole idea of cooperative learning is that students collaborate with each other, so their seating should promote this collaboration.

Optimal in what sense? The main idea. We are looking for the *best (optimal)* placements of students within a group.

Normally, the word "best" is understood in the sense of some *numerical* optimality criterion. However, in our case, it is difficult to formulate the exact *numerical* criterion. What should we do?

Let us borrow from the experience of modern physics and use symmetries. In modern physics, symmetry groups are a tool that enables to compress complicated differential equations into compact form (see, e.g., [14]). Moreover, the very differential equations themselves can be uniquely deduced from the corresponding symmetry requirements (see, e.g., [3, 2]).

It turns out that in many cases, there are reasonable *symmetries*, and it is natural to assume that the (ordinal) optimality criterion is invariant with respect to these symmetries. Then, we are able to describe all choices that are optimal with respect to some invariant ordinal optimality criteria.

This general approach was described and used in [1, 9, 10, 11, 13], in particular, for fuzzy control: to find optimal membership functions, optimal t-norms and t-conorms, and optimal defuzzification procedures. In this section, we will show that this approach is applicable to student placement as well.

We must choose a family of placements. For a given number g of students in a group, we must select a *placement* $P = (\vec{x}_1, \ldots, \vec{x}_g)$, i.e., positions of each student on a planar floor. Of course, the efficiency of a placement depends only on the students' placement with respect to each other, and it should not change if we simply shift or rotate the entire group without changing their relative positions. So, for every placement $P = (\vec{x}_1, \ldots, \vec{x}_g)$, and for every planar motion (rotation + shift) T, the transformed placement $TP = (T(\vec{x}_1), \ldots, T(\vec{x}_g))$ has the exact same learning efficiency. Thus, based on the learning efficiency, we cannot choose a *single* optimal placement P, we can only choose a *family* of all placements $\{TP\}_T$ which correspond to different transformations T.

What is a criterion for choosing a family of placements? We want to select the *best* family of placements. It means that we have some *criterion* that enables us to choose between the two families.

Traditionally, optimality criteria are *numerical*, i.e., to every family F, we assign some value $J(F)$ expressing its quality, and choose a family for which this value is maximal (i.e., when $J(F) \geq J(G)$ for every other alternative G). However, it is not necessary to restrict ourselves to such numeric criteria only.

For example, if we have several different families F that have the same learning ability $A(F)$, we can choose between them the one which leads to the largest class capacity $C(F)$. In this case, the actual criterion that we use to compare two families is not numeric, but more complicated:

> *A family F_1 is better than the family F_2 if and only if*
>
> – *either $A(F_1) > A(F_2)$,*
>
> – *or $A(F_1) = A(F_2)$ and $C(F_1) > C(F_2)$.*

A criterion can be even more complicated.

The only thing that a criterion *must* do is to allow us, for every pair of families (F_1, F_2), to make one of the following conclusions:

- the first family is better with respect to this criterion (we'll denote it by $F_1 \succ F_2$, or $F_2 \prec F_1$);
- with respect to the given criterion, the second family is better ($F_2 \succ F_1$);
- with respect to this criterion, the two families have the same quality (we'll denote it by $F_1 \sim F_2$);
- this criterion does not allow us to compare the two families.

Of course, it is necessary to demand that these choices be consistent.

> For example, if $F_1 \succ F_2$ and $F_2 \succ F_3$ then $F_1 \succ F_3$.

The criterion must be final, i.e., it must pick the unique family as the best one. A natural demand is that this criterion must choose a *unique* optimal family (i.e., a family that is better with respect to this criterion than any other family).

The reason for this demand is very simple: If a criterion *does not choose* any family at all, then it is of no use. If *several* different families are the best according to this criterion, then we still have the problem of choosing the best among them. Therefore we need some additional criterion for that choice, like in the above example:

> If several families $F_1, F_2, \ldots$ turn out to have the same learning ability ($A(F_1) = A(F_2) = \ldots$), we can choose among them a family with largest class capacity ($C(F_i) \to \max$).

So what we actually do in this case is abandon that criterion for which there were several "best" families, and consider a new "composite" criterion instead: F_1 is better than F_2 according to this new criterion if either it was better according to the old criterion, or they had the same quality according to the old criterion and F_1 is better than F_2 according to the additional criterion.

In other words, if a criterion does not allow us to choose a unique best family, it means that this criterion is not final, we'll have to modify it until we come to a final criterion that will have that property.

The criterion should not depend on the enumeration of students. The exact mathematical form of a placement P depends on the exact ordering of students. In principle, we could apply an arbitrary permutation π of the number set $\{1, \ldots, g\}$, and rename the students according to this new permutation, so that a student who was originally assigned # i will now be assigned a new number $\pi(i)$.

It is reasonable to assume that the relative quality of different families should not change under such a permutation, i.e., if the family F is better than a family G, then the transformed family $\tilde{F}$ should also be better than the family $\tilde{G}$.

We are now ready for the formal definitions.

Definition 1. *Let $g > 0$ be a positive integer.*

- *By a placement, we mean a tuple $P = (\vec{x}_1, \ldots, \vec{x}_g)$ of g vectors from R^2.*
- *By a family F, we mean a set $\{TP\}_T$ of all the tuples which can be obtained from some fixed placement P by applying rotations and shifts.*

Denotation. *Let's denote the set of all possible families by Φ. The set of all pairs (F_1, F_2) of elements $F_1 \in \Phi$, $F_2 \in \Phi$, is usually denoted by $\Phi \times \Phi$.*

Definition 2. *An arbitrary subset R of a set of pairs $\Phi \times \Phi$ is called a relation on the set Φ. If $(F_1, F_2) \in R$, it is said that F_1 and F_2 are in relation R; this fact is denoted by $F_1 R F_2$.*

Definition 3. *A pair of relations $(\prec, \sim)$ on a set Φ is called consistent if it satisfies the following conditions, for every $F, G, H \in \Phi$:*

(1) *if $F \prec G$ and $G \prec H$ then $F \prec H$;*
(2) *$F \sim F$;*
(3) *if $F \sim G$ then $G \sim F$;*
(4) *if $F \sim G$ and $G \sim H$ then $F \sim H$;*
(5) *if $F \prec G$ and $G \sim H$ then $F \prec H$;*
(6) *if $F \sim G$ and $G \prec H$ then $F \prec H$;*
(7) *if $F \prec G$ then it is not true that $G \prec F$, and it is not true that $F \sim G$.*

Definition 4. *Assume a set Φ is given. Its elements will be called alternatives.*

- *By an optimality criterion, we mean a consistent pair $(\prec, \sim)$ of relations on the set Φ of all alternatives.*
  - *If $F \succ G$ we say that F is better than G;*
  - *if $F \sim G$ we say that the alternatives F and G are equivalent with respect to this criterion.*
- *We say that an alternative F is optimal (or best) with respect to a criterion $(\prec, \sim)$ if for every other alternative G either $F \succ G$ or $F \sim G$.*

Comment. In this paper, we will consider optimality criteria on the set Φ of all families.

Definition 5. *We say that a criterion is final if there exists an optimal alternative, and this optimal alternative is unique.*

Definition 6. *Let $\pi : \{1, \ldots, g\} \to \{1, \ldots, g\}$ be a permutation.*

- *By a permutation $\pi(P)$ of a placement*
$$P = (\vec{x}_1, \ldots, \vec{x}_g),$$
we mean a placement
$$\pi(P) = (\vec{x}_{\pi(1)}, \ldots, \vec{x}_{\pi(g)}).$$
- *By a permutation $\pi(F)$ of a family of placements F we mean the family consisting of permutations of all placements from F.*

Definition 7. *We say that an optimality criterion on Φ is permutation-invariant if for every two families F and G and for every permutation π, the following two conditions are true:*

i) *if F is better than G in the sense of this criterion (i.e., $F \succ G$), then $\pi(F) \succ \pi(G)$;*
ii) *if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $\pi(F) \sim \pi(G)$.*

Comment. The demands that the optimality criterion is final and permutation-invariant are quite reasonable. At first glance they may seem rather trivial and therefore weak, because these demands do not specify the exact optimality criterion. However, these demands are strong enough, as the following theorem shows:

Theorem. *If a family F is optimal in the sense of some optimality criterion that is final and permutation-invariant, then every placement P from this family F is either a pair ($g = 2$), or an equilateral triangle ($g = 3$).*

Comment. Thus, our general approach provides a precise mathematical justification for the empirical fact that groups of 2 and 3 students are the most efficient.

Proof. This proof is based on the following lemma:

Lemma. *If an optimality criterion is final and permutation-invariant, then the optimal family F_{opt} is also permutation-invariant, i.e., $\pi(F_{opt}) = F_{opt}$ for every permutation π.*

Proof of the Lemma. Since the optimality criterion is final, there exists a unique family F_{opt} that is optimal with respect to this criterion, i.e., for every other F:

- either $F_{opt} \succ F$
- or $F_{opt} \sim F$.

To prove that $F_{opt} = \pi(F_{opt})$, we will first show that the permuted family $\pi(F_{opt})$ is also optimal, i.e., that for every family F:

- either $\pi(F_{opt}) \succ F$
- or $\pi(F_{opt}) \sim F$.

If we prove this optimality, then the desired equality will follow from the fact that our optimality criterion is final and therefore, there is only one optimal family (so, since the families F_{opt} and $\pi(F_{opt})$ are both optimal, they must be the same family).

Let us show that $\pi(F_{opt})$ is indeed optimal. How can we, e.g., prove that $\pi(F_{opt}) \succ F$? Since the optimality criterion is permutation-invariant, the desired relation is equivalent to $F_{opt} \succ \pi^{-1}(F)$, where π^{-1} is the inverse permutation. Similarly, the relation $\pi(F_{opt}) \sim F$ is equivalent to $F_{opt} \sim \pi^{-1}(F)$.

These two equivalences allow us to complete the proof of the proposition. Indeed, since F_{opt} is optimal, we have one of the two possibilities:

- either $F_{opt} \succ \pi^{-1}(F)$,
- or $F_{opt} \sim \pi^{-1}(F)$.

In the first case, we have $\pi(F_{opt}) \succ F$; in the second case, we have $\pi(F_{opt}) \sim F$.

Thus, whatever family F we take, we always have either $\pi(F_{opt}) \succ F$, or $\pi(F_{opt}) \sim F$. Hence, $\pi(F_{opt})$ is indeed optimal and thence, $\pi(F_{opt}) = F_{opt}$. The lemma is proven.

Let us now prove the theorem. Since the criterion is final, there exists an optimal family $F_{opt} = \{TP\}_T$. Due to the Lemma, the optimal family is permutation-invariant. Thus, for every placement P from the optimal family, and for every permutation π, the permutation $\pi(P)$ of this placement also belongs to this optimal family, i.e., $\pi(P) = TP$ for some planar motion T.

Since the distance $d(a, b)$ between the two points does not change under a motion, we can therefore conclude that for every two point $\vec{x}_i$ and $\vec{x}_j$ from the optimal placement P, we have $d\vec{x}_i, \vec{x}_j) = d(\vec{x}_{\pi(i)}, \vec{x}_{\pi(j)})$. In particular, this is true if we pick three different integers i, j, k from 1 to g, and take as π a permutation which permutes i and j and leaves all other integers k unchanged. For this permutation, the above formula leads to $d(\vec{x}_i, \vec{x}_k) = d(\vec{x}_j, \vec{x}_k)$. Thus, for every $i < j$, we have:

- $d(\vec{x}_i, \vec{x}_j) = d(\vec{x}_1, \vec{x}_j)$ and
- $d(\vec{x}_1, \vec{x}_j) = d(\vec{x}_1, \vec{x}_2)$.

Hence,

$$d(\vec{x}_i, \vec{x}_j) = d(\vec{x}_1, \vec{x}_j) = d(\vec{x}_1, \vec{x}_2).$$

So, the distance between every two points from the placement is the same. Therefore:

- If there are three points, we have a unilateral triangle, and
- more than 3 points are (in the plane) impossible

  (because if three points $\vec{x}_1$, $\vec{x}_2$ and $\vec{x}_3$ form a unilateral triangle with a side d, then there is only one point $\vec{x}_4$ which is different from $\vec{x}_)3$ and which has the same distance d from both $\vec{x}_1$ and $\vec{x}_2$, and the distance from this point to $\vec{x}_3$ is different from d).

The theorem is proven.

Acknowledgments. This work was supported in part by NASA under cooperative agreement NCC5-209, by NSF grants No. DUE-9750858 and CDA-9522207, by the United Space Alliance grant No. NAS 9-20000 (PWO C0C67713A6), and by the Future Aerospace Science and Technology Program (FAST) Center for Structural Integrity of Aerospace Systems, effort sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant number F49620-95-1-0518.

The authors are thankful to all participants of the Cooperative Learning Workshop, Cloudcroft, New Mexico, May 18–22, 1998, especially to Andrew P. Bernat, for fruitful discussions and encouragement.

References

[1] B. Bouchon-Meunier *et al.*, "On the formulation of optimization under elastic constraints (with control in mind)", *Fuzzy Sets and Systems,* 1996, Vol. 81, pp. 5–29.

[2] A. Finkelstein, O. Kosheleva, and V. Kreinovich, "Astrogeometry: towards mathematical foundations", *International Journal of Theoretical Physics*, 1997, Vol. 36, No. 4, pp. 1009–1020.

[3] A. M. Finkelstein, V. Kreinovich, and R. R. Zapatrin, "Fundamental physical equations uniquely determined by their symmetry groups," *Lecture Notes in Mathematics*, Springer-Verlag, Berlin-Heidelberg-N.Y., Vol. 1214, 1986, pp. 159–170.

[4] D. W. Johnson, R. T. Johnson, and E. Holubec, *Circles of learning: cooperation in the classroom*, Interactive Book Co., Edina, MN, 1990.

[5] D. W. Johnson, R. T. Johnson, and K. A. Smith, *Active learning: cooperation in the college classroom*, Interactive Book Co., Edina, MN, 1991.

[6] G. Klir and B. Yuan, *Fuzzy sets and fuzzy logic: theory and applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

[7] M. Kochen and A. N. Badre, "On the precision of adjectives which denote fuzzy sets", *J. Cybern.*, 1976, Vol. 4, No. 1, pp. 49–59.

[8] M. Koshelev, "Fuzzy logic explains the golden proportion," *International Journal of Intelligent Systems*, 1997, Vol. 12, pp. 415–417.

[9] V. Kreinovich, C. Quintana, and R. Lea, "What procedure to choose while designing a fuzzy control? Towards mathematical foundations of fuzzy control", *Working Notes of the 1st International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems*, College Station, TX, 1991, pp. 123–130.

[10] V. Kreinovich *et al.*, "What non-linearity to choose? Mathematical foundations of fuzzy control", *Proceedings of the 1992 International Conference on Fuzzy Systems and Intelligent Control*, Louisville, KY, 1992, pp. 349–412.

[11] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

[12] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, CRC Press, Boca Raton, Florida, 1997.

[13] M. H. Smith and V. Kreinovich, "Optimal strategy of switching reasoning methods in fuzzy control", Chapter 6 in H. T. Nguyen, M. Sugeno, R. Tong, and R. Yager (eds.), *Theoretical aspects of fuzzy control*, J. Wiley, N.Y., 1995, pp. 117–146.

[14] *Symmetries in physics*, Springer-Verlag, Berlin, N.Y., 1992.

[15] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes", *IEEE Transactions on Systems, Man and Cybernetics*, 1973, Vol. 3, pp. 28–44.