Decision Making Based on Satellite Images: Optimal Fuzzy Clustering Approach

Vladik Kreinovich[1], Hung T. Nguyen[2], Scott A. Starks[1], and Yeung Yam[3]

[1]NASA Pan American Center for Earth and Environmental Studies
University of Texas, El Paso, TX 79968, USA, email {vladik,sstarks}@utep.edu
[2]Department of Mathematical Sciences, New Mexico State University
Las Cruces, NM 88003, USA, email hunguyen@nmsu.edu
[3]Department of Mechanical & Automation Engineering
The Chinese University of Hong Kong, Shatin, NT, Hong Kong, China
email yyam@mae.cuhk.edu.hk

Abstract

In many real-life decision-making situations, in particular, in processing satellite images, we have an enormous amount of information to process. To speed up the information processing, it is reasonable to first classify the situations into a few meaningful classes (clusters), find the best decision for each class, and then, for each new situation, to apply the decision which is the best for the corresponding class. One of the most efficient clustering methodologies is *fuzzy clustering*, which is based on the use of fuzzy logic. Usually, *heuristic* clusterings are used, i.e., methods which are selected based on their empirical efficiency rather than on their proven optimality. Because of the importance of the corresponding decision making situations, it is therefore desirable to theoretically analyze these empirical choices. In this paper, we formulate the problem of choosing the *optimal* fuzzy clustering as a precise mathematical problem, and we show that in the simplest cases, the empirically best fuzzy clustering methods are indeed optimal.

1 Fuzzy Clustering: Existing Approaches and Formulation of the Problem

For satellite imaging, fuzzy clustering is important. Decision making is especially important in geophysics, because in many geophysical situations, a wrong decision can be very costly (be it digging a well where there is no oil, or not preparing the building for the potential earthquakes, or spending lost of effort on securing building against earthquakes which are not typical for this area). To decrease the possibility of a costly erroneous decision, we must use as much information as possible. One of the important sources of such information is satellite imaging. However, with satellite images, we face a different problem: each satellite image contains a huge amount of data. A good photo contains up to a Gigabyte of information, and with modern multi-spectral satellite images, we get several Gigabytes. We do not know how to process all this information.

One of the known methods of fighting this information explosion is *clustering*. Instead of analyzing each photo individually, we do the following: First, we classify the photos into a few meaningful clusters. Then, for each cluster, we find the best decision. Finally, when we encounter a new situation, we find the cluster to which this situation belongs, and make a decision which is the best for this cluster.

The idea of clustering is very natural in science: The analysis of every new phenomenon starts with *classification*, when instead of *numerous* different *examples*, we have a *few classes*. Classification helped to analyze chemical elements, elementary particles, living organisms, astronomical objects, etc.

In some situations, where assumptions about structure of data can be formulated in statistical terms, statistical techniques (see, e.g., [13]) are appropriate if we have sufficiently many data. In other situations, we must use heuristic classification methods, in particular, methods that use fuzzy logic. The main idea of fuzzy clustering is described in [1, 2, 3, 4, 5, 6, 8, 9, 10, 15, 23, 24].

The goal of fuzzy clustering: "typical" representatives and how to use them. We start with objects which we want to classify (i.e., to cluster). To classify, we use several (numerical) characteristics of these object. Let us denote the total number of these characteristics by s. The s real numbers that characterize each object can be naturally viewed as a point in s-dimensional space R^s. Thus, having n objects means that we have n points $x_1, \ldots, x_n$ in this space. These n points are the input for clustering.

As a result of clustering, we want to describe several clusters. Each cluster can be characterized by its "typical" element $t_j \in R^s$. After these typical elements $t_1, \ldots, t_q$ are found, we can then classify each object $x \in R^s$ according to which typical element it is closest to. This "classification" is a fuzzy notion:

- if an element x is very close to, say, t_1, and not close to any other typical representative, then it is reasonable to conclude that x belongs to class 1;
- however, if an object $x \in R^s$ is almost equally close to two different representatives t_1 and t_2, then it is reasonable to conclude that this object belongs, to some extent, to *both* clusters 1 and 2.

To express this idea in precise terms, we select a function $f(x)$ (called *potential function*) such that for every two point x and y from R^s, the value $f(x-y)$ describes to what extent x and y are close. This function is usually non-negative, and the closer x and y, the larger the value of the potential function. Potentially, as a potential function, we can use a *membership function* which describes the relation "x and y are close"; however, from the mathematical viewpoint, the choice of membership function would mean that we only allow $f(x)$ to take values from the interval $[0,1]$, and sometimes, more general values are needed (in our main text, we will explain why we need such values).

When the potential function is selected, then we can say that an object x belongs to 1-st cluster with a degree $f(x-t_1)$, to the 2-nd cluster with the degree $f(x-t_2)$, $\ldots$, and to q-th cluster with the degree $f(x-t_q)$. Since we do not require any normalization of the function $f(x)$, it is convenient to *normalize* these values so that they will add up to 1, in other words, to describe the degree to which x belongs to j-th cluster as

$$d_j(x) = \frac{f(x-t_j)}{f(x-t_1)+\ldots+f(x-t_q)}. \tag{1}$$

How to find "typical" representatives? The most widely used approach. We have described how to classify an object when the clusters (or, to be more precise, their typical representatives) have already been found. How can we find these representatives?

The most widely used fuzzy clustering method is the method of Fuzzy C-Means (Fuzzy ISODATA) [1, 2, 3, 4, 5, 6, 10, 15]. This method is based on the natural idea that each characteristic of a typical representative should be equal to an average over all elements of the corresponding cluster. If we have *crisp* clustering, then we would simply take the arithmetic average. However, since we have *fuzzy* clustering, it is natural to count, in this average, each element x_i with the weight $d_j(x_i)$ that is proportional to this element's degree of belonging to the cluster. In other words, it is natural to require that for each j,

$$t_j = \frac{d_j(x_1)\cdot x_1 + \ldots + d_j(x_n)\cdot x_n}{d_j(x_1)+\ldots+d_j(x_n)}. \tag{2}$$

This method leads to *good quality* clustering. Its main *disadvantage* is that since the values $d_j(x_i)$, in their turn, depend on t_j, the equation (2) is, actually, a non-linear system of equations for determining the cluster "centers" $t_1, \ldots, t_q$, and solving this system of equations often requires lots of *computation time*.

How to find "typical" representatives? Recent approaches. To simplify computations, a new method has been recently proposed [23, 24] (see also [8, 9]). This method is based on the following idea: when we say that an element t_j is a typical representative of the cluster that consists of elements $x_{i_1}, \ldots, x_{i_k}$, we mean that for each element $x \in R^s$, the degree $f(x-t_j)$ with which x is close to t_j is equal to the average of the degrees $f(x-x_{i_1}), \ldots, f(x-x_{i_k})$ with which x is close to all elements of this cluster:

$$f(x-x_{i_1}) + \ldots + f(x-x_{i_k}) = k\cdot f(x-t_j). \tag{3}$$

If we have a *crisp* classification, then each of the original data points $x_1, \ldots, x_n$ belongs to one and only one cluster and therefore, by adding equalities (3) for all q clusters, we would conclude that

$$\sum_{i=1}^n f(x-x_i) = \sum_{j=1}^q k_j\cdot f(x-t_j), \tag{4}$$

where k_j is the total number of elements in j-th cluster (i.e., the *cardinality* of j-th cluster).

For a *fuzzy* clustering, it is reasonable to expect a similar formula, with k_j being the *fuzzy cardinality* of j-th cluster (see, e.g., [16]). So, to find t_j, we can do the following:

- compute, for all x, the function
$$M(x) = \sum_{i=1}^n f(x-x_i).$$
- represent this function $M(x)$ as a sum
$$M(x) = \sum_{j=1}^q k_j\cdot f(x-t_j)$$
for the smallest possible number of clusters.

Theoretically, the smallest possible number of clusters is 1, in which case $M(x) = k_1\cdot f(x-t_1)$. If one cluster is indeed sufficient, then, due to the properties of the "closeness" function $f(x)$, we can find t_1 easily: it is the value for which $M(x)$ is the largest possible. In this case, if $f(x)$ is normalized in such a way that $f(0)=1$ (i.e., if $f(x)$ is a membership function, and x is close to x with degree of truth 1), we can take $k_1 = M(t_1)$.

In view of this observation, it is reasonable to select, as t_1, the value for which $M(x)$ is the largest possible. In this case, we cannot take $k_1 = M(t_1)$, because other clusters are also contributing to this value $M(t_1)$. Instead, we can take $k_1 = q \cdot M(t_1)$ for some number $q \in (0,1)$. After that, we can subtract $k_1 \cdot f(x-t_1)$ from the original function $M(x)$, and use a similar method to represent the new function $M_1(x) = M(x) - k_1 \cdot f(x-t_1)$ as a sum

$$M_1(x) = \sum_{j=2}^{q} k_j \cdot f(x-t_j);$$

etc. We stop when the remainder becomes small enough.

This method is very similar to a very successful method of image reconstruction used in radio astronomy under the name of *CLEAN* (see, e.g., [14]). Due to the success of the CLEAN method, it is not surprising that this clustering method also turned out to be reasonably successful.

Main problem: how to choose a potential function? We have mentioned that the above fuzzy clustering methods turned out to be very successful, but we must clarify this statement: these methods are very successful provided we appropriately choose the potential function $f(x)$. For a different choice of $f(x)$, the resulting clustering may not be that good.

To the best of our knowledge, so far, the choice of the potential function was mainly done either empirically or heuristically. The following three families of potential functions are most widely used:

- in the original Fuzzy C-Means method, the function $f(x) = |x|^{-m}$ is used, where $|x|$ is the norm of a vector x, and $m > 0$ is a positive real number;
- in [23, 24], the potential function $f(x) = \exp(-\alpha \cdot |x|)$ is used; and
- in [8, 9], the Gaussian potential function $f(x) = \exp(-\alpha \cdot |x|^2)$ is used.

The first choice is used when we have no information about the typical cluster radius; the second and third choices presuppose that an approximate cluster radius is already known.

In this paper, we show that these three choices are indeed optimal in some reasonable sense. Thus, we provide a theoretical justification of these empirical and heuristic choices.

2 Optimal Potential Functions: General Idea

Optimal in what sense? The main idea. We are looking for the *best* (*optimal*) choice of a potential function.

Normally, the word "best" is understood in the sense of some *numerical* optimality criterion. However, in our case of *fuzzy* choice, it is often difficult to formulate the exact *numerical* criterion. Instead, we assume that there is an *ordinal* criterion, i.e., that we can compare arbitrary two choices, but that we cannot assign numerical values to these choices.

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 [7, 18, 19, 20, 21], in particular, for fuzzy control. In this section, we will show that this approach is applicable to fuzzy clustering as well.

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., [22]). Moreover, the very differential equations themselves can be uniquely deduced from the corresponding symmetry requirements (see, e.g., [12, 11]).

It is possible to use symmetry. As we have mentioned, in our previous papers, we have shown that the symmetry group approach can be used to find optimal membership functions, optimal t-norms and t-conorms, and optimal defuzzification procedures.

It is therefore reasonable to expect that the same approach can also be used to choose the best potential function for fuzzy clustering.

3 Optimal Potential Functions: Case When We Do Not Have a Prior Knowledge of the Cluster Radius

We must choose a family of functions. We must select a potential function $f(x)$. The only way the potential function $f(x)$ is used in clustering is through the normalized formula (1). Because of the normalization, if we re-scale the values of the potential function, i.e., if we choose a constant $C > 0$ and consider a new potential function $\tilde{f}(x) = C \cdot f(x)$, this new potential function will lead to exactly the same values $d_j(x)$ as the old one. Therefore, from the viewpoint of fuzzy clustering, there is no way to distinguish between the functions $f(x)$ and $\tilde{f}(x) = C \cdot f(x)$. So, based on clustering behavior, we cannot choose a *single* function $f(x)$; we can only choose a 1-parametric *family* of functions $\{C \cdot f(x)\}$ that is characterized by a parameter C.

Comment about notations. In the following text, we will denote families of functions by capital letters, such as F, F', G, etc.

We must choose the best family of functions. We want to select the *best* family of functions.

What is a criterion for choosing a family of functions? What does it mean to choose a *best* family of functions? 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 classification ability $P(F)$, we can choose between them the one that has the minimal computational complexity $C(F)$. 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 $P(F_1) > P(F_2)$,*
>
> – *or $P(F_1) = P(F_2)$ 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 classification ability ($P(F_1) = P(F_2) = \ldots$), we can choose among them a family with minimal computational complexity ($C(F_i) \to \min$).

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 must not change if we change the measuring unit for x. The exact mathematical form of a function $f(x)$ depends on the exact choice of units for measuring the s coordinates $x^1, \ldots, x^s$ of $x \in R^s$. If we replace each of these units by a new unit that is λ times larger, then the same physical value that was previously described by a numerical value x^k will now be described, in the new units, by a new numerical value $\tilde{x}^k = x^k/\lambda_j$. For example, if we replace centimeters by inches, with $\lambda = 2.54$, then $x^k = 5.08$ cm becomes $\tilde{x}^k = x^k/\lambda = 2$ in. After this transformation, x changes to $\tilde{x} = x/\lambda$.

How will the expression for closeness $f(x)$ change if we use the new units? In terms of $\tilde{x}$, we have $x = \lambda \cdot \tilde{x}$. Thus, if we change the measuring unit for x, the same dynamics that was originally represented by a function $f(x)$, will be described, in the new units, by a function $\tilde{f}(x) = f(\lambda \cdot x)$.

Since we assumed that we have no information about the cluster radii, there is no reason why one choice of unit should be preferable to the other. Therefore, it is reasonable to assume that the relative quality of different families should not change if we simply change the units, 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 $\tilde{G}$.

The criterion must not change if we apply a rotation. Similarly, it is reasonable to require that the relative quality of two different families of functions do not change if we apply an arbitrary *rotation* around 0 in s-dimensional space R^s.

We are now ready for the formal definitions.

Definition 1.

- *By a family F, we mean a differentiable function $f(x)$ from R^s to R.*
- *We say that a function $e(x)$ belongs to the family $f(x)$ (or that $f(x)$ contains the function $e(x)$) if $e(x) = C \cdot f(x)$ for some $C > 0$.*
- *Two families F and G are considered equal if they contain the same functions.*

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 $\Phi \times \Phi$.
- 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 2. *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 3. *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 $F \sim G$.*
- *We say that a criterion is final if there exists an optimal alternative, and this optimal alternative is unique.*

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

Definition 4. *Let $\lambda > 0$ be a positive real number.*

- *By a λ-rescaling of a function $f(x)$ we mean a function $\tilde{f}(x) = f(\lambda \cdot x)$.*
- *By a λ-rescaling of a family of functions F we mean the family consisting of λ-rescalings of all functions from F.*

Denotation. *λ-rescaling of a family F will be denoted by $R_\lambda(F)$.*

Definition 5. *We say that an optimality criterion on Φ is unit-invariant if for every two families F and G and for every number $\lambda > 0$, 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 $R_\lambda(F) \succ R_\lambda(G)$;*

ii) *if F is equivalent to G in the sense of this criterion (i.e., $F \sim G$), then $R_\lambda(F) \sim R_\lambda(G)$.*

Definition 6. *Let $T : R^s \to R^s$ be a rotation around 0 in s-dimensional space.*

- *By a T-rotation of a function $f(x)$ we mean a function $\tilde{f}(x) = f(Tx)$.*
- *By a T-rotation of a family of functions F we mean the family consisting of T-rotations of all functions from F.*

Denotation. *T-rotation of a family F around 0 will be denoted by $T(F)$.*

Definition 7. *We say that an optimality criterion on Φ is rotation-invariant if for every two families F and G and for every rotation T, 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 $T(F) \succ T(G)$;*

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

Comment. As we have already remarked, the demands that the optimality criterion is final, unit-invariant, and rotation 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 1. *If a family F is optimal in the sense of some optimality criterion that is final, unit-invariant, and rotation-invariant, then every function $f(x)$ from this family F has the form $C \cdot |x|^\alpha$ for some real numbers C and α.*

Comments.

- Thus, our general approach provides a precise mathematical justification for the (highly successful) potential functions used in Fuzzy C-Means approach.

- Since none of the optimal functions are from the interval $[0,1]$, our result explains why we cannot restrict ourselves to membership functions $f(x)$, and why we need to consider the potential functions which can attain values outside the interval $[0,1]$.
- The proofs are presented in detail in our Technical Report [17]. For the case when we *have* the prior knowledge of the cluster radius, a similar approach explains the potential functions $f(x)=\exp(-\alpha\cdot|x|)$ and $f(x)=\exp(-\alpha\cdot|x|^2)$.

Acknowledgments. This work was supported in part by NASA under cooperative agreement NCCW-0089, by NSF grants No. DUE-9750858 and EEC-9322370, 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. Part of this work was conducted while one of the author (H.T.N.) was visiting the Department of Mechanical and Automation Engineering at the Chinese University of Hong Kong under the support of RGC Earmarked grant RGC519/95E.

References

[1] J. C. Bezdek, "Numerical taxonomy with fuzzy sets", *Journal of Mathematical Biology*, 1974, Vol. 1, pp. 57–71.

[2] J. C. Bezdek, "Cluster validity with fuzzy sets", *Journal of Cybernetics*, 1973, Vol. 3, No. 3, pp. 58–71.

[3] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum, NY, 1981.

[4] J. C. Bezdek, R. Hathaway, M. Sabin, and W. Tucker, "Convergence theory for fuzzy C-Means: counterexample and repairs", *IEEE Trans. Systems, Man, and Cybernetics*, 1987, Vol. SMC-17, pp. 873–877.

[5] J. C. Bezdek, R. Hathaway, M. Sabin, and W. Tucker, "Convergence theory for fuzzy C-Means: counterexample and repairs", In: J. Bezdek (ed.), *The Analysis of Fuzzy Information*, CRC Press, 1987, Vol. 3, Chapter 8.

[6] J. C. Bezdek and S. K. Pal (eds.) *Fuzzy models for pattern recognition*, IEEE Press, N.Y., 1992.

[7] 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.

[8] S. Chiu, "Fuzzy model identification based on cluster estimation", *J. of Intelligent and Fuzzy Systems*, 1994, Vol. 2, No. 3, pp. 267–278.

[9] S. Chiu, "Selecting input variables for fuzzy models", *J. of Intelligent and Fuzzy Systems*, 1996, Vol. 4, pp. 243–256.

[10] J. C. Dunn, "A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters", *Journal of Cybernetics*, 1973, Vol. 3, pp. 32–57.

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

[12] 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.

[13] K. Fukunaga, *Introduction to statistical pattern recognition*, Academic Press, San Diego, CA, 1990.

[14] *Galactic and extra-galactic radio astronomy*, Springer-Verlag, NY, 1974.

[15] A. Kandel, *Fuzzy techniques in pattern recognition*, Wiley-Interscience, NY, 1982.

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

[17] V. Kreinovich, H. T. Nguyen, and Y. Yam, "Optimal Choices of Potential Functions in Fuzzy Clustering", *The Chinese University of Hong Kong, Department of Mechanical & Automation Engineering*, Technical Report CUHK-MAE-98-001, January 1998.

[18] 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.

[19] 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.

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

[21] 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.

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

[23] R. R. Yager and D. P. Filev, "Approximate clustering via the mountain method", *IEEE Trans. Systems, Man and Cybernetics*, 1994, Vol. 24, pp. 1279–1284.

[24] R. R. Yager and D. P. Filev, "Generation of fuzzy rules by mountain clustering", *Journal of Intelligent and Fuzzy Systems*, 1994, Vol. 2, No. 3, pp. 209–219.