MAXIMUM ENTROPY APPROACH TO FUZZY CONTROL

Arthur Ramer
School of CSE, University of New South Wales,
P.O. Box 1, Kensington, N,S.W. 2033, Australia

Vladik Kreinovich
Computer Science Department, University of Texas at El Paso, El Paso, TX 79968
vladik@cs.ep.utexas.edu

1. INTRODUCTION

Traditional control theory helps only in the situations when we know the behavior of the system that we are going to control. However, in many situations, e.g., in space exploration, we do not have this knowledge, but we still have to make control decisions. In such situations, a reasonable idea is to find an operator who is good in this kind of control, and translate his control experience into a precise formula. This control experience is not usually formulated in terms of the natural-language rules like "if x is small, then control must also be small". The methodology of translating these rules into an actual control strategy was proposed by L. Zadeh under the name of *fuzzy control* [CZ72, M74] (for a survey see, e.g., [L90]).

Three steps are necessary to specify this translation: first, we must determine membership functions that correspond to all natural- language terms (like "small" or "big") that appear in the rules. Second, we must choose operations that correspond to & and $\vee$. As a result we get a membership function $\pi_C(u)$ for a control; then we need a method to transform this function $\pi_C(u)$ into a single control value (a *defuzzification method*).

Different choices lead to control strategies of drastically different quality, so which alternative to choose is very important [KQL91, KQLFLKBR92].

In the present paper, we analyze two possible choices:

1) First, we consider the problems of choosing $\&-$ and $\vee-$operations. When we make a choice, we thus restrict the set of possible control strategies. Since a wrong choice can lead to a low quality control, it sounds reasonable to try to loose as few possibilities as possible. In other words, it sounds reasonable to choose $\&-$ and $\vee-$operations in such a way that the uncertainty corresponding to $\pi_C(u)$ is the biggest possible. This methodology is well known in the case when the uncertainty is probabilistic; it is called a *maximum entropy approach*, and it is widely applied to various problems ranging from processing physical data to processing uncertainties in expert systems [J79, KK79, C83, K89].

   Just like in a probabilistic case, we want to evaluate the uncertainty of a membership function as the average number of binary questions that one needs to ask in order to determine the value. In the present paper, we propose the formulas that compute this uncertainty (in Section 2), and find out for what operations this uncertainty is the biggest (in Section 3). We prove that the desired maximal uncertainty is attained when we use $\min(a+b,1)$ for $\vee$, and $\min(a,b)$ for &.

In control terms, this maximum entropy approach is proved to lead to maximally stable controls. This result is in good accordance with common sense: we minimized the lost opportunities, and therefore, we ended up with the best possible control.

2) The above arguments are reasonable only if we are ready to apply various defuzzification techniques to extract the best control from $\pi_C(u)$. However, in industrial applications, a defuzzification rule is usually fixed. Since this rule is not necessarily the most appropriate (see., e.g., [YPL92, KQLFLKBR92]), it is reasonable to try to depend on it to a smallest extent. In other words, in these cases, it is reasonable to choose & and $\vee-$ operations from the condition that the uncertainty related to $\pi_C(u)$ is the *smallest* possible. In the present paper, we find the operations for which the resulting uncertainty is the smallest possible (in Section 4).

   We prove that this leads to the choice of $\max(a, b)$ for $\vee$ and ab for &. In control terms, this minimum entropy approach leads to maximally smooth controls. This result is also in good corresponds with common sense: since we were extremely cautious, we ended up with a very smooth control.

2. HOW TO MEASURE UNCERTAINTY THAT CORRESPONDS TO A MEMBERSHIP FUNCTION

General motivation

In order to answer this question, let us recall where the values $\pi(x)$ of a membership function come from. If $\pi(x)$ corresponds to, say, "small", then $\pi(x)$ is our degree of belief that x is small. One of the most natural ways to express this "degree of belief" by a number is to ask several experts, whether they consider x small or not, and after M out of N answer "yes", take M/N as $\pi(x)$ (see, e.g., [KF88]). This approach allows us to interpret the value $\pi(x)$ as a frequency (or, if you like, *subjective probability*) that x is small.

With this interpretation in mind, let us estimate the uncertainty $U(\pi)$ that corresponds to a function $\pi(x)$. Suppose that we have a notion (like "small") that is described by a function $\pi(x)$. If the only thing we know about some real value x is that it satisfies this property (e.g., "is small"), then how many binary questions do we have to ask to determine x?

To make the definitions easier to understand, we will proceed as follows:

- first, we start with the case of finitely many alternatives, when it is easy to estimate uncertainty;
- then, we move on to more and more complicated membership functions,
- and, finally, we figure out how to describe uncertainty for a general membership function.

We want our definitions to be natural and understandable, therefore, before each definition, there is a big motivation part. Readers who are interested only in the precise mathematical definitions can skip these motivations.

Preliminary case:
how to define uncertainty in discrete case.

We want to estimate uncertainty by the number of binary questions that we have to ask to get the complete knowledge. This number is easy to define and to understand in the discrete case, when we have finitely many alternatives $x_1, ..., x_n$. In this case, it is sufficient to ask finitely many questions to determine the actual alternative. Therefore, our uncertainty can be estimated as the smallest number of binary ("yes-no") questions that we have to ask to determine x_i. Let us briefly remind how.

If we ask q questions, then we have 2^q different combinations of answers. Therefore, after we know these answers, we can distinguish between no more than 2^q different alternatives. On the other hand, if we have 2^q alternatives, then we can use the standard bisection method to find the right alternative after q questions. So, if we have $n = 2^q$ alternatives, then the smallest number of questions that we can use is $q = \log_2(n)$. If we have n alternatives, where $2^{q-1} < n < 2^q$, then we also need q questions. So, in general, for n alternatives, we need $Q = \lceil \log_2(n) \rceil$ questions.

Let us give a formal definition. **Definition 1.** By a *number of alternatives* n we will mean a positive integer. If the number of alternatives is equal to n, then by a *number of questions, necessary to determine an alternative*, we will mean the value $\lceil \log_2(n) \rceil$.

Comment. If we have only one situation with n alternatives, then we have to ask all Q questions. But suppose that we have several (say N) situations, in each of which there are n alternatives, and we want to determine a proper alternative in each of these situations. Totally, we have $n \times n \times ... \times n = n^N$ combinations of alternatives. Therefore, to determine the appropriate combination, it is necessary to ask $\lceil \log_2(n^N) \rceil$ questions. So, for N situations, we need to ask $Q(N) = \lceil \log_2(n^N) \rceil$ binary questions. By an average number of questions, we will mean the ratio $Q(N)/N$. As an uncertainty, let us take the limit of this average number of questions when $N \to \infty$. Let us give formal definitions.

Definition 2. By a *number of alternatives* n we will mean a positive integer. If $N > 0$, then by an *average number of questions that are necessary to determine an alternative* we mean a ratio $Q(N)/N$, where $Q(N) = \lceil \log_2(n^N) \rceil$. The limit $\lim_{N \to \infty} Q(N)/N$ will be called the *uncertainty* (or *entropy*) that corresponds to this number of alternatives.

PROPOSITION 1. *The uncertainty that corresponds to n alternatives is equal to* $\log_2(n)$.

(In order to avoid disrupting the reader's attention, we placed all the proofs in a special Section 5.)

Now, let us start describing uncertainty for membership functions, starting with the simplest ones.

Case 1: $\pi(x)$ is a crisp interval

We want to start with the cases when the fuzzy set is crisp. Before we do that, let us recall the basic definitions.

Definition 3. By a *membership function* we mean a function $\pi : R \to [0,1]$ that is not identically 0. If for every x, $\pi(x) \in \{0,1\}$, then this function is called *crisp.* A crisp membership function will also be called a *crisp set.* For each set $S \subset R$, we can define its *characteristic function* $\pi_S(x)$ as follows: $\pi_S(x) = 1$ if $x \in S$, and $\pi_S(x) = 0$ if $x \notin S$.

Informal formulation of a problem. Let us first consider the case, when a membership function $\pi(x)$ is a characteristic function of an interval $[a,b]$, i.e., it is equal to 1 inside the interval, and to 0 outside it.

In this case, if x satisfies the property that is expressed by this membership function, it simply means that $x \in [a,b]$. How to estimate the amount of uncertainty in this statement?

How to estimate uncertainty if the set of alternatives is infinite: an idea. In our case, the set of all alternatives coincides with the set of all points on an interval and is, therefore, infinite. If we ask finitely many questions, then we will have only finitely many possible answers. Therefore, to find the precise value of x, we need to ask infinitely many questions. What to do?

The natural solution of this problem is related to the fact that in real life, we never know precise values of physical quantities. We can measure them, but all the measurements have finite precision. The only information that we can get from a measurement procedure is a value $\tilde{x}$ such that the actual value x satisfies the inequality $|x - \tilde{x}| \le \varepsilon$, where $\varepsilon > 0$ is the precision of this measurement. In other words, the actual value x belongs to the interval $[\tilde{x} - \varepsilon, \tilde{x} + \varepsilon]$. After we know such an $\tilde{x}$, we say that we know x with precision ε.

Therefore, to estimate of the information that $x \in [a,b]$, let us fix some ε, and estimate the number of binary questions that we have to ask in order to find x with a precision ε.

Now, we are ready to give a formal definition.

Definition 4. Let S be a set of real numbers, and $\varepsilon > 0$. By a *possible set of* $\varepsilon-$*alternatives* we mean a set $\{x_1, ..., x_n\}$ of real numbers such that

$$S \subset \bigcup_{i=1}^{n} [x_i - \varepsilon, x_i + \varepsilon].$$

By an *uncertainty* of this set of alternatives $\{x_i\}$ we mean the value $\log_2(n)$.

Comments.

1. These values are called $\varepsilon-$*alternatives* because if we know which of the values x_i is $\varepsilon-$close to the actual values x, then we know x with a precision ε. They are called *alternatives*, because according to the above definition, for each possible x (i.e., for each $x \in S$) there exists an x_i, for which $|x - x_i| \le \varepsilon$.

2. For those readers who are knowledgeable in approximation theory, the following remark will be helpful: The notion that we have just introduced, turns out to be a particular case of a well-known mathematical notion: namely, an arbitrary possible set of $\varepsilon-$alternatives is a $\varepsilon-$net for S (see, e.g., [L66]).

3. For one and the same set S, we can have several different sets of $\varepsilon-$alternatives. Different sets of $\varepsilon-$alternatives may have different number of elements, and thus lead to different values of uncertainty. As a measure of uncertainty of S, it is natural to consider the smallest of these values. So, we arrive at the following definition:

Definition 5. Let S be a set of real numbers, and $\varepsilon > 0$. By an $\varepsilon-$*uncertainty* $Q(\pi_S, \varepsilon)$ of a set S, we mean the uncertainty of the smallest possible set of $\varepsilon-$alternatives.

Comments.

1. This denotation may look somewhat clumsy: why did we denote it $Q(\pi_S, \varepsilon)$, and not simply $Q(S, \varepsilon)$? The reason is that later on, we will define $Q(\pi, \varepsilon)$ for membership functions π that do not necessarily describe crisp sets.

2. The value of $Q(\pi_S, \varepsilon)$ is easy to estimate when S is an interval $[a, b]$:

PROPOSITION 2. $Q(\pi_{[a,b]}, \varepsilon) = \log_2(\lceil (b-a)/(2\varepsilon) \rceil)$.

Comments.

1. The precise proof is placed in Section 5. It may look too mathematical, but its idea is very simple: To determine where x is with the precision ε, we must divide the entire interval $[a, b]$ of length $b-a$ into $N \approx (b-a)/(2\varepsilon)$ subintervals of length 2ε, and determine in which of them x is (for that, we need $\approx \log_2(N)$ questions).

2. By explicitly mentioning a precision ε when we described uncertainty, we ended up with a finite value of uncertainty. But the drawback is that instead of one value that described uncertainty in a finite case, we now have a function (i.e., infinitely many values) that correspond to different $\varepsilon > 0$. Functions are more difficult to compare, more difficult to store and analyze. So, it is desirable to describe uncertainty by one number (or a few numbers).

To do that, let us remark that for big ε (to be more precise, for $\varepsilon \geq (b-a)/2$) one possible alternative $x_1 = (a+b)/2$ is sufficient, hence $Q(\pi_{[a,b]}, \varepsilon) = \log_2(1) = 0$. So, what matters is the behavior of $Q(\pi, \varepsilon)$ for small $\varepsilon > 0$. And for small ε, the above expression for $Q(\pi, \varepsilon)$ has a simple asymptotics:

Definition 6. We say that two functions $f(\varepsilon)$ and $g(\varepsilon)$ are *asymptotically equivalent*, and denote it by $f(\varepsilon) \sim g(\varepsilon)$ if $\lim_{\varepsilon \to 0} f(\varepsilon) = \infty$, and $\lim_{\varepsilon \to 0}(f(\varepsilon) - g(\varepsilon)) = 0$.

PROPOSITION 3. *When* $\varepsilon \to 0$, *then* $Q(\pi_{[a,b]}, \varepsilon) \sim \log_2(b-a) - \log_2(2\varepsilon)$.

Comment. From this Proposition it follows that when we consider intervals, the behavior of the function $Q(\pi_{[a,b]}, \varepsilon)$ for small $\varepsilon > 0$ is uniquely determined by a single number $\log_2(b-a)$. We will see that the same is true for more complicated sets. As we will see, for fuzzy (non-crisp) membership functions π, we will need two numbers to describe the behavior of $Q(\pi, \varepsilon)$: the coefficients at 1 and $\log_2(2\varepsilon)$. Therefore, we can take these coefficients as a measure of uncertainty that corresponds to a membership function π. Namely, we arrive at the following Definition.

Definition 7. Let π be a membership function. We say that for π, *uncertainty is defined* if $Q(\pi, \varepsilon) \sim u - m \log_2(2\varepsilon)$ for some real numbers u and m. The pair (u, m) is called the *uncertainty* of a fuzzy set π. The value of u will be denoted by $U(\pi)$, and the value of m by $m(\pi)$.

From Proposition 3, we immediately get the following corollary:

COROLLARY 1. *For an interval* $[a, b]$*,* $U(\pi_{[a,b]} = \log_2(b - a)$ *and* $m(\pi_{[a,b]}) = 1$*.*

Remark. To avoid possible misunderstanding, we need to make the following remark. Up to now we defined uncertainty as a number (or an average number) of binary questions that we need to ask to determine an alternative (uniquely or with the precision ε). The number of questions is always non-negative. Therefore, all previously defined measures of uncertainty were non-negative.

However, as one can easily see, the expression $U(\pi_{[a,b]}) = \log_2(b - a)$ *can* be negative if $b - a < 1$. There is not contradiction in there, because this number $U(\pi_{[a,b]})$ is *not* equal to the (non-negative) number of questions. To get a number of questions $Q(\pi_{[a,b]}, \varepsilon)$, one has to restrict oneself to small ε, and add to this number $U(\pi_{[a,b]})$ the term $-\log_2(2\varepsilon)$. For small ε, this second term is positive, and moreover, it tends to $+\infty$ as $\varepsilon \to 0$. Therefore, for sufficiently small ε, the resulting sum is positive.

For the readers who know the definition of entropy of a random variable, the following analogy will be helpful. For discrete variables, that take only finitely many values x_i with probabilities p_i, the entropy is usually defined as $-\sum_i p_i \log_2(p_i)$. This sum is always non-negative. But if we have a continuous variable, with a probability density $\rho(x)$, then we cannot apply this formula. For continuous variables, the usual analog of entropy is the integral $S = -\int \rho(x) \log_2(\rho(x))\, dx$. Unlike discrete entropy, this integral can be negative: for example, if we take a uniform distribution on an interval $[a, b]$, i.e., a function $f(x) = 1/(b - a)$ for $x \in [a, b]$ and 0 outside this interval, then this entropy is equal to $\log_2(b - a)$, and this value is negative for $b - a < 1$. The reason is the same as for our non-statistical definition: Shannon's entropy $-\sum_i p_i \log_2(p_i)$ can be interpreted as an average number of questions, and therefore, is always non-negative. But its continuous analog is only indirectly related to the number of questions: namely, $S - \log_2(2\varepsilon)$ is the average number of questions that we have to ask to determine x with precision ε.

Case 2: $\pi(x)$ is a general crisp set

Definition 8. We say that a set S is a *union of disjoint intervals* if

$$S = \bigcup_{i=1}^{n} [a_i, b_i]$$

for some n, $\{a_i\}$ and $\{b_i\}$ such that $a_i \leq b_i$ for all i, $a_i < b_i$ for some i, and the intervals $[a_i, b_i]$ and $[a_j, b_j]$ have no common points if $i \neq j$.

Comment. For each such set S, its Lebesgue measure (total length) is easily computable as $\mu(S) = \sum_{i=1}^{n}(b_i - a_i)$.

PROPOSITION 4. *For an arbitrary set S that is a union of disjoint intervals, uncertainty is defined for π_S, $U(\pi_S) = \log_2(\mu(S))$, and $m(\pi_S) = 1$.*

Case 3: $\pi(x)$ is piecewise constant

Definition 9. We say that a membership function $\pi(x)$ is *normalized* if $\sup_x \pi(x) = 1$.

Definition 10. We say that a membership function is *piecewise constant* if there exist values $x_1 < x_2 < ... < x_n$ such that $\pi(x) = 0$ for $x < x_1$ and $x > x_n$, $\pi(x) = const$ on each of the intervals (x_i, x_{i+1}), and for each i, $\pi(x_i)$ coincides either with the value of $\pi(x)$ for $x < x_i$, or with the values of $\pi(x)$ for $x > x_i$.

How to estimate the number of binary questions when the membership function is normalized. If a function is piecewise constant, then it takes only finitely many different values. Let us order them in a sequence $h_0 = 0 < h_1 < h_2 < ... < h_k$. For such functions, $\sup \pi(x) = h_k$. In this section, we consider only normalized functions, so we can assume that $h_k = 1$.

Let us first consider an example. Suppose that a membership function $\pi(x)$ is equal to 1 for $x \in [-a, a]$, is equal to 0.6 for $x \in [-2a, -a]$ and $x \in [a, 2a]$, and $\pi(x) = 0$ for $x \notin [-2a, 2a]$. In this case, 60% of the experts believe that the area of possible values of x is the interval $[-a, a]$ of length $2a$, and the remaining 40% believe that x is in the interval $[-2a, 2a]$ (of length $4a$). If the experts from this 60% majority are right, then we need $\sim \log_2(2a) - \log_2 \varepsilon$ binary questions. If the minority experts are right, then we need $\sim \log_2(4a) - \log_2 \varepsilon$ questions. Since all the experts are considered equally good, it is reasonable to assume that in general, in 60% of the cases the majority is right, and in 40% of cases, the minority is right. Therefore, the average number of binary questions that we have to ask in order to locate x in an interval of length ε, is $\sim 0.6(\log_2(2a) - \log_2 \varepsilon) + 0.4(\log_2(4a) - \log_2 \varepsilon) = (0.6 \log_2(2a) + 0.4 \log_2(4a)) - \log_2 \varepsilon$.

In the general case, the h_1-th part of all the experts believe that x belongs to the set $\{x : \pi(x) \geq h_1\}$, $(h_2 - h_1)$ of them believe that $x \in \{x : \pi(x) \geq h_2\}$, $(h_3 - h_2)$ of them believe that $x \in \{x : \pi(x) \geq h_3\}$, ..., and $h_k - h_{k-1}$ of them believe that $x \in \{x : \pi(x) \geq h_k\}$. If $x \in \{x : \pi(x) \geq h_1\}$, then we need $Q(\{x : \pi(x) \geq h_1\}, \varepsilon)$ questions to determine x with the precision ε. If $x \in \{x : \pi(x) \geq h_2\}$, then we need $Q(\{x : \pi(x) \geq h_2\}, \varepsilon)$ questions, etc.

Therefore, according to the opinion of h_1 of experts, we need $Q(\{x : \pi(x) \geq h_1\}, \varepsilon)$ questions; according to the opinion of $(h_2 - h_1)$ of the experts, we need to ask $Q(\{x : \pi(x) \geq h_2\}, \varepsilon)$ questions, etc.

So, it is natural to define the expected number of questions as

$$Q(\pi, \varepsilon) = \sum_{i=0}^{k-1} (h_{i+1} - h_i) Q(\pi_{\{x : \pi(x) \geq h_{i+1}\}}, \varepsilon).$$

How to estimate the number of binary questions for a non-normalized membership function. Suppose, for example, that some term from the natural language is represented by a (fuzzy) membership function $\pi(x)$ that is equal to 0.6 for all x from $[a, b]$, and to 0 for all other x. In view of the above interpretation of $\pi(x)$, it means that only 60% of all the experts had any opinion about what the initial natural language term means, and the others simply gave no answers. In this case, we can take into consideration only the opinions of those who said something.

As we have already argued, it is reasonable to assume that in 60% of cases the majority is right. Therefore, the only thing that we know about the average number of questions is that it is $\geq 0.6Q(\pi_{[a,b]}, \varepsilon) \sim 0.6(\log_2(b - a) - \log_2(2\varepsilon))$. When the remaining 40% of the experts make their decisions, there may be more questions, but right now, when we are given the above-described membership function $\pi(x)$, this number $0.6Q(\pi_{[a,b]}, \varepsilon)$ is the only estimate that we can get. So, it is reasonable to take it as the description of uncertainty of this membership function $\pi(x)$.

In the general case, according to the opinion of h_1 of experts, we need $Q(\{x : \pi(x) \geq h_1\}, \varepsilon)$ questions; according to the opinion of $(h_2 - h_1)$ of the experts, we need to ask $Q(\{x : \pi(x) \geq h_2\}, \varepsilon)$ questions, etc.

So, it is reasonable define the expected number of questions as

$$Q(\pi, \varepsilon) = \sum_{i=0}^{k-1} (h_{i+1} - h_i) Q(\pi_{\{x:\pi(x) \geq h_{i+1}\}}, \varepsilon).$$

In both cases, we justified the following definition:

Definition 11. Let $\pi(x)$ be a piecewise-constant membership function, that takes only the values $h_0 = 0 < h_1 < ... < h_k$. By its $\varepsilon-$*uncertainty* $Q(\pi, \varepsilon)$ we mean a value

$$Q(\pi, \varepsilon) = \sum_{i=0}^{k-1} (h_{i+1} - h_i) Q(\pi_{\{x:\pi(x) \geq h_{i+1}\}}, \varepsilon).$$

PROPOSITION 5. *For an arbitrary piecewise constant normalized membership function* $\pi(x)$*, uncertainty* $(U(\pi), m(\pi))$ *is defined,* $m(\pi) = 1$ *and*

$$U(\pi) = \sum_{i=0}^{k-1} (h_{i+1} - h_i) \log_2(\mu\{x : \pi(x) \geq h_{i+1}\}).$$

PROPOSITION 6. *For an arbitrary piecewise constant normalized membership function* $\pi(x)$*, uncertainty* $(U(\pi), m(\pi))$ *is defined,* $m(\pi) = \sup_x \pi(x)$*, and*

$$U(\pi) = \sum_{i=0}^{k-1} (h_{i+1} - h_i) \log_2(\mu\{x : \pi(x) \geq h_{i+1}\}).$$

General case: an arbitrary membership function $\pi(x)$ that can appear in control problems.

Let us rule out monster functions, and consider only regular ones. Before we start our definitions, we want to make one remark. Initially, we defined a membership function as an arbitrary mapping from R to [0,1]. From mathematical viewpoint, a function can be non-measurable, or exotic in other senses. There may be some reason in analyzing them, and estimating their uncertainty. But the main objective of the present paper is to analyze uncertainty with respect to fuzzy control problems. And membership functions, that are usually considered in fuzzy control (see, e.g., [L90]), are all rather regular. Namely, their domain can be divided into finitely many intervals of monotonicity (i.e., intervals, on which this function is either increasing or decreasing).

So, although our definitions can in principle be applied to more exotic functions as well, let us not sink deep into mathematics, and let us restrict ourselves only to functions that can actually appear in fuzzy control problems. So, let us give the following definition.

Definition 12. We say that a membership function $\pi(x)$ is *regular* if $\pi(x) \to 0$ as $|x| \to \infty$, and there exist values $x_1 < x_2 < ... < x_n$ such that on each of the intervals $(-\infty, x_1)$, (x_i, x_{i+1}), $1 \leq i \leq n-1$, and (x_n, ∞), the function $\pi(x)$ is either monotonic increasing or monotonic decreasing, and for each i, the value $\pi(x_i)$ coincide either with the left-side limit $\lim_{\varepsilon>0,\varepsilon\to 0} \pi(x-\varepsilon)$, or with the right-side limit $\lim_{\varepsilon>0,\varepsilon\to 0} \pi(x+\varepsilon)$.

Comment. Since $\pi(x) \to 0$ for $x \to -\infty$, and $\pi(x) \geq 0$ for all x, a regular function $\pi(x)$ can only be increasing on $(-\infty, x_1)$. Likewise, it can only be decreasing on $[x_n, \infty)$.

The idea of the following definition is a follows. When we define a membership function in mathematical terms, we say that its values $\pi(x)$ are real numbers. To describe a real number with infinite precision, one needs infinitely many bits. Therefore, even if we have a procedure that would allow us to describe the experts' degree of belief with a better and better precision, after any finite number of questions, we can have only an approximate value of $\pi(x)$.

The reader may notice the similarity between this argument and the one for Case 1: there, we argued that we always know x with some precision; here, we use similar arguments to convince the reader that we can know $\pi(x)$ only with some precision.

On each stage of determining $\pi(x)$, we only know it with some precision δ. So, on this stage, there are only finitely many distinguishable degrees of belief $h_0 = 0 < h_1 < h_2 < ... < h_k = 1$ such that $|h_{i+1} - h_i| \leq \delta$. In other words, all the values of $\pi(x)$ from $h_0 = 0$ to h_1 are indistinguishable from 0, all the values from h_1 to h_2 are indistinguishable from h_1, etc. If we change the values $\pi(x)$ to the corresponding values h_i, then we get a piecewise constant function $\bar{\pi}(x)$ that is (on this stage) quite possible. Therefore, the uncertainty $U(\bar{m})$ of this membership function is a possible (on this stage) value of uncertainty $U(\pi)$.

Since on each stage, many such functions $\bar{m}$ can be possible, on each stage, we get the entire interval of the possible values of $U(\bar{m})$. After the next measurement, the set of possible functions $\bar{m}$ diminishes, hence the resulting interval becomes smaller. If for $\delta \to 0$ this interval shrinks to a point, then it is natural to call the value that corresponds to this point the uncertainty of the limit membership function π.

Let us now give a formal definition.

Definition 13. By a *monotonic sequence* we mean a sequence $\vec{h} = (h_0, h_1, ..., h_k)$ of real numbers such that $h_0 = 0 < h_1 < h_2 < ... < h_k = 1$. We say that this sequence is $\delta-$*precise* (where δ is some positive real number) if $h_{i+1} - h_i \le \delta$ for all i.

Definition 14. For an arbitrary monotonic sequence $\vec{h}$, we define a *projection* $pr_{\vec{h}} : [0,1] \to [0,1]$ as follows: $pr_{\vec{h}}(y) = 0$ for $y < h_1$ and $pr_{\vec{h}}(y) = h_i$ for $h_i \le y < h_{i+1}$.

PROPOSITION 7. *If $\pi(x)$ is a membership function, and $\vec{h}$ is any sequence, then $\bar{\pi}(x) = pr_{\vec{v}}(\pi(x))$ is a piecewise constant membership function.*

Comment. Due to this Proposition, for each of these "projected" membership functions $\bar{\pi}(x) = pr_{\vec{v}}(\pi(x))$, uncertainty $U(\bar{\pi})$ is well defined.

Definition 15. Let $\delta > 0$, and $\pi(x)$ be a regular membership function. We say that a pair of real numbers (u, m) is a $\delta-$*possible* value of uncertainty for π, if $u = U(pr_{\vec{h}}(\pi(x))$ and $m = m(pr_{\vec{h}}(\pi(x))$ for some $\delta-$precise sequence $\vec{h}$.

Definition 16. Let $\pi(x)$ be a regular membership function. We say that its *uncertainty is defined and is equal to* (u, m) if the following is true: for arbitrary sequences $\delta_n \to 0$ and (u_n, m_n), if (u_n, m_n) is a δ_n-possible value of uncertainty for π, then $(u_n, m_n) \to (u, m)$. These values u and m will be denoted by $U(\pi)$ and $m(\pi)$.

Main result of this Section: how to estimate uncertainty

THEOREM 1. *For an arbitrary regular membership function, uncertainty is defined, $m(\pi) = \sup_x \pi(x)$, and*

$$U(\pi) = \int_0^{\sup_x \pi(x)} \log_2(\mu(\{x : \pi(x) \ge h\}))\, dh.$$

COROLLARY 2. *For a normalized regular function $\pi(x)$, $m(\pi) = 1$ and*

$$U(\pi) = \int_0^1 \log_2(\mu(\{x : \pi(x) \ge h\}))\, dh.$$

For a monotonic normalized function $\pi(x)$, this expression can be further simplified:

THEOREM 2. *Let $x_0 > 0$ be a positive real number. Assume that a function $\pi(x)$ is such that $\pi(x) = 0$ for $x < 0$, $\pi(0) = 1$, $\pi(x) = 0$ for $x \ge x_0$, and for $x \in [0, x_0)$, $\pi(x)$ is continuous and decreasing. Then,*

$$U(\pi) = \log_2(x_0) - \frac{1}{\ln(2)} \int_0^{x_0} \frac{1 - \pi(x)}{x}\, dx.$$

Comment. The first term in this formula for $U(\pi)$ is the uncertainty that corresponds to just belonging to $[0, x_0]$. So, we can interpret this Theorem as follows: If we know only that $x \in [0, x_0]$, than our uncertainty in the value of x can be (according to Corollary 1) estimated as $\log_2(x_0)$. If we add the additional information that x belongs to a fuzzy set that is described by a membership function $\pi(x)$, then we diminish our certainty to the value $U(\pi) = \log_2(x_0) - (1/\ln(2)) \int_0^{x_0} (1 - \pi(x))/x\, dx$. This decrease in uncertainty $U(\pi_{[0,x_0]}) - U(\pi)$ measures the information that is brought by this additional knowledge that x belongs to a fuzzy set. So, as a numerical value of the *information* $I(\pi)$ that is contained in this knowledge, we can take this difference in uncertainties $(1/\ln(2)) \int_0^{x_0} (1 - \pi(x))/x\, dx$.

It is interesting to mention that the resulting expression for the information practically coincides with the expression $I(\pi) = \int_0^{x_0} (1-\pi(x))/x\, dx$ that was introduced (from different assumptions) in [R90, RY92]. The factor $1/\ln(2)$ that makes them different is simply due to the fact that we use $\log_2$ for entropy, while some other authors use natural logarithms.

How to compare uncertainties of different membership functions?

Comment. In the following text, we will look for operations that lead to the biggest and to the smallest uncertainty. To formalize that, we must learn to compare uncertainties of different membership functions.

When we use entropy $S(p)$ as a definition of uncertainty of a probability distribution p, we just say that the uncertainty of a distribution p_1 is bigger than the uncertainty of p_2 if $S(p_1) \geq S(p_2)$.

But we have defined uncertainty as pair of numbers, so we must figure out when one pair is bigger than the other. For that, we can use the fact that we defined uncertainty $(U(\pi), m(\pi))$ as such a pair that $Q(\pi, \varepsilon) \sim U(\pi) - m(\pi) \log_2(2\varepsilon)$, and we have already argued that what matters is the behavior of Q when $\varepsilon \to 0$. So, we can use the following definition:

Definition 17. We say that uncertainty (u_1, m_1) is *bigger* than the uncertainty (u_2, m_2), and denote it by $(u_1, m_1) \succ (u_2, m_2)$ if there exists an ε_0 such that for all $\varepsilon < \varepsilon_0$, $u_1 - m_1 \log_2(2\varepsilon) > u_2 - m_2 \log_2(2\varepsilon)$.

PROPOSITION 8. *$(u_1, m_1) \succ (u_2, m_2)$ if and only if either $m_1 > m_2$, or $m_1 = m_2$ and $u_1 > u_2$.*

Comment. So, $\succ$ is the lexicographic order on the set of all pairs of real numbers.

Denotation 1. We will use the standard denotation $(u_1, m_1) \succeq (u_2, m_2)$ to express that the uncertainty (u_1, m_1) is either bigger than (u_2, m_2), or equal to (u_2, m_2).

3. OPERATIONS FOR WHICH UNCERTAINTY IS THE BIGGEST

What is a fuzzy control? A brief mathematical explanation

Definition 18. Assume that a set $S \subset R^n$ is given. Elements $\vec{x} = (x_1, ..., x_n) \in S$ will be called *states.*

Comment. Informally, the values $x_1, ..., x_n$ describe everything that we need to know to make a control decision. For example, if we control a heater/cooler, then $n = 1$, and the only variable we need to know is the difference $x_1 = t - t_0$ between the actual and the desired temperature. If we are controlling a spaceship, then we need to know its coordinates x_1, x_2, x_3, its current velocity vector (3 more variables $x_4 = \dot{x}_1, x_5 = \dot{x}_2, x_6 = \dot{x}_3$), and 2 angles that describe the orientation. So, for a spaceship, $n = 8$.

Definition 19. Let us fix a finite set $\mathcal{P}$ of continuous membership functions. The elements of this set will be called *fuzzy properties.*

Comment. As examples, we can consider "small", "big", "medium", etc.

Definition 20. By an *elementary formula* E we mean an expression of the type $P_i(x_i)$, where P_i is a fuzzy property. By a *rule*, we mean an expression of the type $E_1, ..., E_m \to P(u)$, where E_i are elementary formulas, P is a fuzzy property, and u is a special variable reserved for control. Formulas E_i are called *conditions*, $P(u)$ is called a *conclusion* of the rule. By a *knowledge base* we mean a finite set of rules.

Comment. As an example of the rule, one can consider a rule $N(x_1) \to N(u)$, meaning that if the difference $t - t_0$ between the actual and the desired temperatures is negligible, then the control should be negligible. Another possible rule is: $SP(x_1) \to SN(u)$, meaning that if the difference $t - t_0$ is small positive, then we need to apply a small negative control (i.e., switch on the cooler a little bit). A similar rule $SN(x_1) \to SP(u)$ tells that if it becomes a little bit cold, it is necessary to switch on the heater for a while.

Motivation of the following definitions. If we have a set of rules, then we can say that a control u is appropriate if and only if one the rules is applicable, and u appropriate according to this rule. Let us denote the statement "control u is appropriate" by $C(u)$. Then, for the three rules that describe the cooler/heater, we have the following informal "formula" that describes when a control u is appropriate:

$$C(u) \equiv (N(x)\&N(u)) \vee (SP(x)\&SN(u)) \vee (SN(x)\&SP(u)).$$

Since $N(x)$, $N(u)$, ..., are fuzzy statements, we can get only fuzzy conclusions about the control, i.e., thus defined $C(u)$ also becomes a fuzzy statement. To get the precise values, we need to choose some operations that would describe & and $\vee$ for fuzzy values.

Initially, L. Zadeh proposed to use min and max, but he stressed that these operations "are not the only operations in terms of which the union and intersection can be defined", and "which of these ... definitions is more appropriate depends on the context" [Z75, pp. 225–226]. So, we will use the maximally general operations.

After we choose operations that correspond to & and $\vee$, we can use the above formula to describe for each u, what is the reasonable degree of belief that this value u is an appropriate control. In other words, we will be able to generate a membership function $\pi_C(u)$ that corresponds to control.

After that, we need some *defuzzification* procedure that would transform this membership function into a single recommended control value. We are interested in comparing different functions $\pi_C(u)$, so we will not describe defuzzification procedures here (their description can be found in the above-cited surveys on fuzzy control).

So, to give the precise definitions, we need to know what $\&-$ and $\vee-$ operations can be.

The meaning of the $\&-$operation (we will denote it by $f_\&(a,b)$) is as follows: Suppose that we have two statements A and B. Our degree of belief in A is equal to a, and our degree of belief in B is equal to b. If we have no other information about A and B, what must the reasonable degree of belief in $A\&B$ equal to? This reasonable degree of belief will be denoted by $f_\&(a,b)$.

In the same situation, a reasonable degree of believe in $A \vee B$ will be denoted by $f_\vee(a,b)$, and $f_\vee$ will be called an $\vee-$operation.

In describing uncertainty of a membership function, we used the interpretation of membership values $\pi(x)$ as frequencies. Namely, we assumed that as a truth value $t(A)$ of an uncertain statement A, we take the ratio $t(A) = N(A)/N$, where $N(A)$ is the number of experts who believe in A, and N is the total number of experts that were questioned. In this interpretation, the following inequalities are true: $N(A \vee B) \le N(A) + N(B)$, $N(A \vee B) \le N$, $N(A \vee B) \ge N(A)$ and $N(A \vee B) \ge N(B)$. If we divide both sides of these inequalities by N, and combine them into one, we get the following inequality: $\max(t(A), t(B)) \le t(A \vee B) \le \min(t(A) + t(B), 1)$, hence $\max(a,b) \le f_\vee(a,b) \le \min(a+b, 1)$. Likewise, from $N(A\&B) \le N(A)$ and $N(A\&B) \le N(B)$ we conclude that $t(A\&B) \le \min(t(A), t(B))$ and $f_\&(a,b) \le \min(a,b)$.

If belief in A and belief in B were independent events, then we would have $t(A\&B) = t(A)t(B)$. In real life, beliefs are not independent: if an expert has strong beliefs in several statements that later turn out to be true, then this means that he is really a good expert, and therefore it is reasonable to expect that his degree of belief in other statements that are true is bigger. If A and B are complicated statements, then many of those experts who believe in A are really good experts, and therefore they believe in B as well (and hence in $A\&B$). Therefore, the total number $N(A\&B)$ of experts who believe in $A\&B$ must be bigger than the same number in the case when beliefs in A and B were uncorrelated random events. So we come to a conclusion that the following inequality sounds reasonable: $t(A\&B) \ge t(A)t(B)$, i.e., $f_\&(a,b) \ge ab$. In statistical terms we can express this inequality by saying that A and B are *non-negatively correlated.*

So, we arrive at the following definitions:

Definition 21. By an *and-or pair* we will understand a pair of continuous functions

$f_\&, f_\vee : [0,1] \times [0,1] \to [0,1]$, that are non-decreasing in both variables, and satisfy the following conditions:

- $\max(a,b) \le f_\&(a,b) \le \min(a+b,1)$;
- $f_\vee(a,b) \le \min(a,b)$;
- $f_\&(0,a) = 0, f_\&(1,a) = a, f_\vee(0,a) = a, f_\vee(1,a) = 1$;
- $f_\vee(a,b) = f_\vee(b,a), f_\&(a,b) = f_\&(b,a)$.

An and-or pair is called *correlated* if $f_\&(a,b) \ge ab$ for all a and b.

Comment. The motivations for this definition are simple:

- If A is false, then $A\&B$ is also false, so $f_\&(0,a) = 0$ for all a.
- If A is true, then $A\&B$ is true if and only if B is true, so in this case $t(A\&B) = t(B)$ hence $f_\&(a,1) = a$ for all a.
- When we say that A and B are both true or that B and A are both true, we mean the same thing. Therefore, $t(A\&B)$ must be always equal to $t(B\&A)$, or, in other words, $f_\&(a,b) = f_\&(b,a)$ for all a, b.
- If our degree of belief in A increases, then our degree of belief in $A\&B$ becomes greater or equal (but cannot become smaller). So the function $f_\&$ must be non-decreasing in both variables.
- If our degrees of belief in A and B change a little bit, then our degree of belief in $A\&B$ cannot change essentially. The smaller is the change in $t(A)$, $t(B)$, the smaller must be the change in $t(A\&B)$. In other words, the function $f_\&$ must be continuous.

Similar arguments justify the condition on $f_\vee$.

Denotation 2. For three numbers a, b, c, we define $f_\&(a,b,c) = f_\&(f_\&(a,b),c)$. For more than three numbers $a, b, ..., c$, we define $f_\&(a,b,...,c) = f_\&(...(f_\&(f_\&(a,b),...)c)$. Likewise, we define $f_\vee(a,b,...,c) = f_\vee(...(f_\vee(f_\vee(a,b),...)c)$.

Comment. The idea of this denotation is simple: the statement $A\&B\&C$ is usually understood as $(A\&B)\&C$, and $A \vee B \vee C$ is understood as $(A \vee B) \vee C$.

Now, we can define fuzzy control.

Definition 22. Suppose that we are given a knowledge base $K = \{R_1, R_2, ...\}$, an and-or pair $(f_\&, f_\vee)$, and a state $\vec{x} \in S$. By a *membership function, that corresponds to a rule* $P_1(x_{i_1}), ..., P_m(x_{i_m}) \to P(u)$, we mean a function $\pi_R = f_\&(P_1(x_{i_1}), ..., P_m(x_{i_m}, P(u))$. By a *control membership function, that corresponds to the knowledge base and a state* $\vec{x}$, we mean a function $\pi_C(u) = f_\vee(\pi_{R_1}, \pi_{R_2}, ...)$, where $R_1, R_2, ...$ are all the rules from K.

Why should we consider control that corresponds to the biggest uncertainty?

When we make a choice of $\&-$ and $\vee-$operations, we thus restrict the set of possible control strategies. Since a wrong choice can lead to a low quality control, it sounds reasonable to try to loose as few possibilities as possible. In other words, it sounds reasonable to choose $\&-$ and $\vee-$operations in such a way that the uncertainty corresponding to $\pi_C(u)$ is the biggest possible.

This methodology is well known in the case when the uncertainty is probabilistic; it is called a *maximum entropy approach*, and it is widely applied to various problems ranging from processing physical data to processing uncertainties in expert systems [J79, KK79, C83, K89].

Result

THEOREM 3. *Suppose that K is a knowledge base, and $\vec{x}$ is a state. Let us denote by $\pi_C(u)$ the control membership function, that corresponds to an arbitrary and- or pair $(f_\&(a,b), f_\vee(a,b))$, and by $\tilde{\pi}_C(u)$ the control membership function that corresponds to the and-or pair $(\min(a,b), \min(a+b,1))$. Then, $(U(\tilde{\pi}_C), m(\tilde{\pi}_C)) \succeq (U(\pi_C), m(\pi_C))$.*

Comment. In other words, the biggest uncertainty is attained when we use $\min(a,b)$ for &, and $\min(a+b,1)$ for $\vee$.

It has been proved in [KQLFLKBR92] that these very and-or operations lead to maximally stable controls. This result is in good accordance with common sense: we minimized the lost opportunities, and therefore, we ended up with the best possible control.

4. OPERATIONS FOR WHICH UNCERTAINTY IS THE SMALLEST

Reasons for choosing operations that lead to the smallest uncertainty. The arguments that we gave in Section 3 are reasonable only if we are ready to apply various defuzzification techniques to extract the best control from $\pi_C(u)$. However, in industrial applications, a defuzzification rule is usually fixed. Since this rule is not necessarily the most appropriate (see., e.g., [YPL92, KQLFLKBR92]), it is reasonable to try to depend on it to a smallest extent. In other words, in these cases, it is reasonable to choose & and $\vee-$ operations from the condition that the uncertainty related to $\pi_C(u)$ is the *smallest* possible. For this problem, we prove the following result.

THEOREM 4. *Suppose that K is a knowledge base, and $\vec{x}$ is a state. Let us denote by $\pi_C(u)$ the control membership function, that corresponds to an arbitrary correlated and-or pair $(f_\&(a,b), f_\vee(a,b))$, and by $\tilde{\pi}_C(u)$ the control membership function that corresponds to the and-or pair $(ab, \max(a,b)$. Then, $(U(\pi_C), m(\pi_C)) \succeq (U(\tilde{\pi}_C), m(\tilde{\pi}_C))$.*

Comment. In other words, the smallest uncertainty is attained when we use ab for &, and $\max(a,b)$ for $\vee$.

It has been proved in [KQLFLKBR92] that these very and-or operations lead to maximally smooth controls. This result is also in good corresponds with common sense: since we were extremely cautious, we ended up with a very smooth control.

5. PROOFS

Proof of Proposition 1. For an arbitrary real number x, $x \le \lceil x \rceil < x+1$, therefore, $|x - \lceil x \rceil| \le 1$. In particular, $|\log_2(n^N) - \lceil \log_2(n^N) \rceil| = |\log_2(n^N) - Q(N)| \le 1$. But $\log_2(n^N) = N \log_2(n)$, so $|Q(N) - N\log_2(n)| \le 1$. Dividing both sides of this inequality by N, we conclude that $|Q(N)/N - \log_2(n)| \le 1/N$, hence $\lim_{N\to\infty} Q(N)/N = \log_2(n)$. Q.E.D.

Proof of Proposition 2. If $\{x_1, ..., x_n\}$ is the possible set of $\varepsilon-$alternatives, then, according to Definition 4, the entire interval $[a,b]$ must be covered by n intervals $[x_i - \varepsilon, x_i + \varepsilon]$

of length 2ε. Therefore, the length $b-a$ of the interval $[a,b]$ must not exceed the sum of their lengths, i.e., $2n\varepsilon$. So, $2n\varepsilon \geq b-a$, hence $n \geq (b-a)/(2\varepsilon)$. Since n is an integer, we have $n \geq \lceil (b-a)/(2\varepsilon) \rceil$.

On the other hand, if we take $x_1 = a+\varepsilon$, $x_2 = a+3\varepsilon$, ..., $x_i = a+(2i-1)\varepsilon$, ..., then we get a possible set of $\varepsilon-$alternatives that contains precisely $N = \lceil (b-a)/(2\varepsilon) \rceil$ elements.

Hence, N is the smallest possible number of elements in a set of possible $\varepsilon-$alternatives, so $Q([a,b],\varepsilon) = \log_2(N)$. Q.E.D.

Proof of Proposition 3. We must prove that $|\log_2(N) - \log_2((b-a)/(2\varepsilon))| \to 0$ when $\varepsilon \to 0$, where $N = \lceil (b-a)/(2\varepsilon) \rceil$. If we denote $\Delta = \lceil (b-a)/(2\varepsilon) \rceil - N$, then this difference can be rewritten as follows:

$$\log_2((b-a)/(2\varepsilon)) - \log_2(N) = \log_2(N+\Delta) - \log_2(N) = \log_2((N+\Delta)/N) = \log_2(1+\Delta/N).$$

But $|x - \lceil x \rceil| \leq 1$ for an arbitrary real number x, in particular, $|\Delta| \leq 1$, since $\Delta = \lceil (b-a)/(2\varepsilon) \rceil - N$. So, $(\Delta/N) \leq 1/N \to 0$, hence $\log_2(1+\Delta/N) \to \log_2(1+0) = 0$. Q.E.D.

Proof of Proposition 4 is similar to Proposition 2.

Proof of Propositions 5 and 6. For each i, due to Proposition 4, we have

$$Q(\pi_{\{x:\pi(x)\geq h_{i+1}\}},\varepsilon) \sim \log_2(\mu\{x:\pi(x)\geq h_{i+1}\}) - \log_2(2\varepsilon).$$

Therefore,

$$Q(\pi,\varepsilon) = \sum_{i=0}^{k-1}(h_{i+1}-h_i)Q(\pi_{\{x:\pi(x)\geq h_{i+1}\}},\varepsilon) \sim$$

$$\sum_{i=0}^{k-1}(h_{i+1}-h_i)\log_2(\mu\{x:\pi(x)\geq h_{i+1}\}) - (\sum_{i=0}^{k-1}(h_{i+1}-h_i))\log_2(2\varepsilon).$$

But

$$\sum_{i=0}^{k-1}(h_{i+1}-h_i) = (h_1-h_0)+(h_2-h_1)+...+(h_k-h_{k-1}) = h_k - h_0 = \sup_x \pi(x) - 0 = \sup_x \pi(x),$$

therefore,

$$Q(\pi,\varepsilon) \sim \sum_{i=0}^{k-1}(h_{i+1}-h_i)\log_2(\mu\{x:\pi(x)\geq h_{i+1}\}) - (\sup_x \pi(x))\log_2(2\varepsilon).$$

So, we get the desired formulas for $U(\pi)$ and $m(\pi)$. Q.E.D.

Proof of Proposition 7. Let $\bar{\pi}(x)$ be a projection of π on $\vec{h} = (h_0, h_1, ..., h_k)$. Then, the function $\bar{\pi}(x)$ takes only the values $h_0, ..., h_j$, where j is the biggest value for which

$h_j \leq \sup_x \pi(x)$. According to the definition of a projection, for each of these values h_i, the inequality $\bar{\pi}(x) \geq h_i$ is equivalent to $\pi(x) \geq h_i$. Therefore, the set $\{x : \pi(x) \geq h_i\}$ coincides with the set $\{x : \pi(x) \geq h_i\}$. On each interval I of monotonicity, the set $\{x \in I : \pi(x) \geq h_i\}$ is an interval. Therefore, the entire set $\{x : \pi(x) \geq h_i\}$ is equal to the finite union of the intervals $\{x \in I_k : \pi(x) \geq h_i\}$ that correspond to the different intervals of monotonicity I_k.

Therefore, the set $\{x : \bar{\pi}(x) \geq h_i\}$ can be represented as a union of intervals. Since the function $\bar{\pi}(x)$ takes only finitely many values $h_0, h_1, \ldots$, the value of $\bar{\pi}(x)$ is equal to h_i if and only if it is $\geq h_i$ and not $\geq h_{i+1}$. So, the level set $\{x : \bar{\pi}(x) = h_i\}$ coincides with the difference between the two sets $\{x : \bar{\pi}(x) \geq h_i\}$ and $\{x : \pi(x) \geq h_{i+1}\}$. Both these sets are finite unions of intervals, therefore, their union can also be represented as a finite union of intervals.

So, the function $\bar{\pi}(x)$ takes only finitely many values h_i, and for each of them, the level set $\{x : \bar{\pi}(x) = h_i\}$ is set is the finite union of intervals. So, the function $\bar{\pi}(x)$ is piecewise constant. Q.E.D.

Proof of Theorem 1. According to Proposition 7, for an arbitrary projection $\bar{\pi}(x)$, $m(\bar{\pi}) = h_j$, where j is the biggest value for which $h_j \leq \sup_x \pi(x)$, and

$$U(\bar{m}) = \sum_{i=0}^{j} (h_{i+1} - h_i) \log_2(\mu\{x : \bar{\pi}(x) \geq h_{i+1}\}).$$

Here, as we have already noticed in the proof of Proposition 7, $\bar{\pi}(x) \geq h_{i+1}$ if and only if $\pi(x) \geq h_{i+1}$. Therefore,

$$U(\bar{m}) = \sum_{i=0}^{j} (h_{i+1} - h_i) \log_2(\mu\{x : \pi(x) \geq h_{i+1}\}).$$

When $\delta \to 0$, $h_j \to \sup_x \pi(x)$. The sum $\sum_{i=0}^{j} (h_{i+1} - h_i) \log_2(\mu\{x : \pi(x) \geq h_{i+1}\})$ is an integral sum for the integral $\int_0^{\sup_x \pi(x)} \log_2(\mu\{x : \pi(x) \geq h\})\, dh$, and therefore, tends to this integral as $\delta \to 0$. Q.E.D.

Proof of Theorem 2. Let us start with an expression from Conclusion 2. Since the function $\pi(x)$ is decreasing for $x > 0$, we get $\{x : \pi(x) \geq h\} = [0, \pi^{-1}(x)]$, where $\pi^{-1}(x)$ denotes an inverse function. Therefore, $\mu(\{x : \pi(x) \geq h\}) = \pi^{-1}(x)$. So, $U(\pi) = \int_0^1 dh \log_2(\pi^{-1}(h))$. Let us introduce the new variable $x = \pi^{-1}(h)$. When $h = 0$, we have $x = x_0$; when $h = 1$, we have $x = 0$; $h = \pi(x)$, so $dh = d(\pi(x))$. Therefore, $U(\pi) = \int_{x_0}^0 d\pi(x)\, log_2(x)$.

We would like to compute this integral using integration by parts. We cannot apply it directly to this expression, because $\pi(x) \log_2(x)$ is infinite for $x = 0$. We can avoid this infinity, if we take into consideration that $(1 - \pi(x)) \log_2(x)$ already tends to 0 as $x \to 0$, and $d\pi(x) = d(\pi(x) - 1) = -d(1 - \pi(x))$. Therefore, $U(\pi) = -\int_{x_0}^0 d(1 - \pi(x))\, log_2(x) =$

$\int_0^{x_0} d(1-\pi(x))\, log_2(x) = (1-\pi(x))\log_2(x)|_0^{x_0} - \int_0^{x_0}(1-\pi(x))d(log_2(x))$. For $x = 0$, $(1-\pi(x))\log_2(x) = 0$, and for $x = x_0$, it is equal to $(1-\pi(x_0))\log_2(x_0)$. Since $\pi(x_0) = 0$, this term is equal to $\log_2(x_0)$.

Now, $\log_2(x) = \ln(x)/\ln(2)$, therefore, $d(\log_2(x)) = (dx/x)/\ln(2)$. So,

$$U(\pi) = (1-\pi(x))\log_2(x)|_0^{x_0} - \int_0^{x_0}(1-\pi(x))d(log_2(x)) =$$

$$\log_2(x_0) - \frac{1}{\ln(2)}\int_0^{x_0}\frac{1-\pi(x)}{x}\,dx.$$

Q.E.D.

Proof of Proposition 8.

1) Let us first prove that if either $m_1 > m_2$, or $m_1 = m_2$ and $u_1 > u_2$, then $u_1 - m_1\log_2(2\varepsilon) > u_2 - m_2\log_2(2\varepsilon)$ for sufficiently small ε.

Indeed, if $m_1 = m_2$ and $u_1 > u_2$, then this inequality is evidently true for all ε. If $m_1 > m_2$, then this inequality is equivalent to $(m_1 - m_2)(-\log_2(2\varepsilon)) > u_2 - u_1$. The left-hand side tends to $+\infty$ as $\varepsilon \to 0$, therefore for sufficiently small ε it is bigger than the right-hand side.

2) To complete the proof, it is now sufficient to prove that if it is not true that either $m_1 > m_2$, or $m_1 = m_2$ and $u_1 > u_2$, then it is not true that $u_1 - m_1\log_2(2\varepsilon) > u_2 - m_2\log_2(2\varepsilon)$ for sufficiently small ε.

Indeed, when can it be not true that either $m_1 > m_2$, or $m_1 = m_2$ and $u_1 > u_2$? In one of the three cases: either when $m_1 = m_2$ and $u_1 = u_2$, or when $m_1 < m_2$, or when $m_1 = m_2$ and $u_1 < u_2$. In the first case the expressions $u_1 - m_1\log_2(2\varepsilon)$ and $u_2 - m_2\log_2(2\varepsilon)$ coincide for all ε, so the inequality is impossible. In the second and third cases, using 1), we can conclude that the opposite inequality $u_2 - m_2\log_2(2\varepsilon) > u_1 - m_1\log_2(2\varepsilon)$ is true for sufficiently small ε. Q.E.D.

Proof of Theorem 3. This proof, as well as the proof of Theorem 4, will be based on the following Lemma:

LEMMA. *If $\pi_1(x)$ and $\pi_2(x)$ are membership functions, and $\pi_1(x) \ge \pi_2(x)$ for all x, then $(U(\pi_1), m(\pi_1)) \succeq (U(\pi_2), m(\pi_2))$.*

Proof of the Lemma. First, since $\pi_1(x) \ge \pi_2(x)$ for all x, we can conclude that $m(\pi_1) = \sup_x \pi_1(x) \ge \sup_x \pi_2(x) = m(\pi_2)$. If $m(\pi_1) > m(\pi_2)$, then $(U(\pi_1), m(\pi_1)) \succeq (U(\pi_2), m(\pi_2))$.

So, to prove the Lemma, it remains to prove that if $m(\pi_1) = m(\pi_2)$, then $U(\pi_1) \ge U(\pi_2)$. Indeed, since we assumed that $\pi_1(x) \ge \pi_2(x)$, if $\pi_2(x) \ge h$, then $\pi_1(x) \ge h$. So,

$$\{x : \pi_2(x) \ge h\} \subset \{x : \pi_1(x) \ge h\}.$$

Therefore, for every h, $\mu(\{x : \pi_1(x) \geq h\}) \geq \mu(\{x : \pi_2(x) \geq h\})$. Since we assumed that $m(\pi_1) = \sup_x \pi_1(x) = \sup_x \pi_2(x) = m(\pi_2)$, we can integrate this inequality from 0 to $m(\pi_i) = \sup_x \pi_i(x)$, and conclude that

$$\int_0^{\sup_x \pi_1(x)} \mu(\{x : \pi_1(x) \geq h\})\, dh \geq \int_0^{\sup_x \pi_2(x)} \mu(\{x : \pi_2(x) \geq h\})\, dh,$$

i.e., that $U(\pi_1) \geq U(\pi_2)$. Q.E.D.

Now, let us prove Theorem 3. We will prove that $\tilde{\pi}_C(u) \geq \pi_C(u)$ and then use the Lemma.

According to Definition 22, $\pi_C(u) = f_\vee(\pi_{R_1}, \pi_{R_2}, ...)$, where $R_1, R_2, ...$ are all the rules form the knowledge base K, and for each rule R of the type $P_1(x_{i_1}), ..., P_m(x_{i_m}) \rightarrow P(u)$, we define π_R as $\pi_R = f_\&(P_1(x_{i_1}), ..., P_m(x_{i_m}), P(u))$. Likewise, $\tilde{\pi}_C(u) = \tilde{f}_\vee(\tilde{\pi}_{R_1}, \tilde{\pi}_{R_2}, ...)$, where $\tilde{f}_\vee(a, b) = \min(a + b, 1)$, and $\tilde{\pi}_R = \min(P_1(x_{i_1}), ..., P_m(x_{i_m}, P(u))$.

1^o. According to our Definition 21 (of an and-or pair), $f_\&(a, b) \leq \min(a, b)$ for all a, b. Therefore, for each rule R_k, $\min(P_1(x_{i_1}), ..., P_m(x_{i_m}), P(u)) \geq f_\&(P_1(x_{i_1}), ..., P_m(x_{i_m}), P(u))$, i.e., $\tilde{\pi}_{R_k} \geq \pi_{R_k}$.

2^o. According to Definition 21, $f_\vee$ if non- decreasing in each of the variables. So, from $\tilde{\pi}_{R_k} \geq \pi_{R_k}$, we conclude that $\pi_C(u) = f_\vee(\pi_{R_1}, \pi_{R_2}, ...) \leq f_\vee(\tilde{\pi}_{R_1}, \tilde{\pi}_{R_2}, ...)$.

3^o. Now, according to Definition 21, $f_\vee(a, b) \leq \tilde{f}_\vee(a, b)$ for all a, b. Hence, $f_\vee(\tilde{\pi}_{R_1}, \tilde{\pi}_{R_2}, ...) \leq \tilde{f}_\vee(\tilde{\pi}_{R_1}, \tilde{\pi}_{R_2}, ...) = \tilde{\pi}_C(u)$.

Combining the inequalities 2^o and 3^o, we conclude that $\pi_C(u) \leq \tilde{\pi}_C(u)$. Application of the Lemma completes the proof of Theorem 3. Q.E.D.

Proof of Theorem 4. This proof is similar to the proof of Theorem 3, and uses the same Lemma. So, to prove that $(U(\pi_C), m(\pi_C)) \succeq (U(\tilde{\pi}_C), m(\tilde{\pi}_C))$, we will prove that $\tilde{\pi}_C(u) \leq \pi_C(u)$, and then use the Lemma.

Here, $\pi_C(u) = f_\vee(\pi_{R_1}, \pi_{R_2}, ...)$, where and for each rule R of the type $P_1(x_{i_1}), ..., P_m(x_{i_m}) \rightarrow P(u)$, $\pi_R = f_\&(P_1(x_{i_1}), ..., P_m(x_{i_m}), P(u))$. Likewise, $\tilde{\pi}_C(u) = \max(\tilde{\pi}_{R_1}, \tilde{\pi}_{R_2}, ...)$, where $\tilde{\pi}_R = \tilde{f}_\&(P_1(x_{i_1}), ..., P_m(x_{i_m}), P(u))$ and $\tilde{f}_\&(a, b) = ab$.

1^o. According to our definition of a correlated and-or pair, $f_\&(a, b) \geq ab = \tilde{f}_\&(a, b)$ for all a, b. Therefore, for every rule R_k, $f_\&(P_1(x_{i_1}), ..., P_m(x_{i_m}), P(u)) \geq \tilde{f}_\&(P_1(x_{i_1}), ..., P_m(x_{i_m}), P(u))$, i.e., $\pi_{R_k} \geq \tilde{\pi}_{R_k}$.

2^o. According to Definition 21, $f_\vee$ if non- decreasing in each of the variables. So, from $\pi_{R_k} \geq \pi_{R_k}$, we conclude that $\pi_C(u) = f_\vee(\pi_{R_1}, \pi_{R_2}, ...) \geq f_\vee(\tilde{\pi}_{R_1}, \tilde{\pi}_{R_2}, ...)$.

3^o. Now, according to Definition 21, $f_\vee(a, b) \geq \max(a, b)$ for all a, b. Hence, $f_\vee(\tilde{\pi}_{R_1}, \tilde{\pi}_{R_2}, ...) \geq \max(\tilde{\pi}_{R_1}, \tilde{\pi}_{R_2}, ...) = \tilde{\pi}_C(u)$.

Combining the inequalities 2^o and 3^o, we conclude that $\pi_C(u) \geq \tilde{\pi}_C(u)$. Application of the Lemma completes the proof of Theorem 4. Q.E.D.

ACKNOWLEDGEMENTS. This work was supported by NSF Grant No. CDA – 9015006, NASA Research Grant No. 9-482 and the Institute for Manufacturing and Materials Management grant. The authors are thankful to Bob Lea, Ron Yager, and John Yen for encouraging discussions.

REFERENCES

[CZ72] S. S. L. Chang and L. A. Zadeh. *On fuzzy mapping and control*, IEEE Transactions on Systems, Man and Cybernetics, 1972, Vol. SMC-2, pp. 30–34.

[C83] P. Cheeseman. *In defense of probability* in Proceedings of the 8-th International Joint Conference on AI, Los Angeles, 1972, pp. 1002–1009.

[J79] E. T. Jaynes. *Where do we stand on maximum entropy?* in R. D. Levine and M. Tribus (Eds.) *The maximum entropy formalism*, MIT Press, Cambridge, MA, 1979.

[KF88] G. J. Klir and T. A. Folger. *Fuzzy sets, uncertainty and information.* Prentice Hall, Englewood Cliffs, NJ, 1988.

[KK79] O. M. Kosheleva, V. Kreinovich. *A letter on maximum entropy method.* Nature, 1979, Vol. 281, pp. 708–709.

[K89] V. Kreinovich. *Entropy approach for the description of uncertainty in knowledge bases*, Technical Report, Center for the New Informational Technology "Informatika", Leningrad, 1989 (in Russian).

[KQL91] 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.

[KQLFLKBR92] V. Kreinovich, C. Quintana, R. Lea, O. Fuentes, A. Lokshin, S. Kumar, I. Boricheva, and L. Reznik. *What non-linearity to choose? Mathematical foundations of fuzzy control.* Proceedings of the 1992 International Fuzzy Systems and Intelligent Control Conference, Louisville, KY, 1992, pp. 349–412.

[L66] G. G. Lorentz. *Approximation of Functions*, Holt, Rinehart and Winston, N.Y., 1966.

[L90] C. C. Lee. *Fuzzy logic in control systems: fuzzy logic controller.* IEEE Transactions on Systems, Man and Cybernetics, 1990, Vol. 20, No. 2, pp. 404–435.

[M74] E. H. Mamdani. *Application of fuzzy algorithms for control of simple dynamic plant*, Proceedings of the IEE, 1974, Vol. 121, No. 12, pp. 1585–1588.

[R90] A. Ramer. *Concepts of fuzzy information measures on continuous domains*, Int. J. General Systems, 1990, Vol. 17, pp. 241–248.

[RY92] A. Ramer, R. Yager. *Analysis of specificity of fuzzy sets*, Proceedings of the IEEE International Conference on Fuzzy Systems, San Diego, CA, 1992, pp. 1097–1104.

[YPL92] J. Yen, N. Pfluger, and R. Langari. *A defuzzification strategy for a fuzzy logic controller employing prohibitive information in command formulation*, Proceedings of IEEE International Conference on Fuzzy Systems, San Diego, CA, March 1992.

[Z75] L. A. Zadeh. *The concept of a linguistic variable and its application to approximate reasoning*, Part 1, Information Sciences, 1975, Vol. 8, pp. 199–249.