Editorial: Fuzzy Models – What Are They, and Why?

(J.C. Bezdek, IEEE Transactions on Fuzzy Systems, Vol. 1, No. 1, February 1993 – Edited by P.D.)

Fuzzy sets are a generalization of conventional set theory that were introduced by Zadeh in 1965 as a mathematical way to represent vagueness in everyday life [1]. The basic idea of fuzzy sets is easy to grasp. Suppose, as you approach a red light, you must advise a driving student when to apply the brakes. Would you say, "Begin braking *74 feet* from the crosswalk"? Or would your advice be more like, "Apply the brakes *pretty soon*"? The latter, of course; the former instruction is too precise to be implemented. This illustrates that precision may be quite useless, while vague directions can be interpreted and acted upon. Everyday language is one example of ways vagueness is used and propagated. Children quickly learn how to interpret and implement fuzzy instructions ("go to bed *about 10*"). We all assimilate and use (act on) fuzzy data, vague rules, and imprecise information, just as we are able to make decisions about situations which seem to be governed by an element of chance. Accordingly, computational models of real systems should also be able to recognize, represent, manipulate, interpret, and use (act on) both fuzzy and statistical uncertainties.

Fuzzy interpretations of data structures are a very natural and intuitively plausible way to formulate and solve various problems. Conventional (crisp) sets contain objects that satisfy *precise properties* required for membership. The set of numbers H from 6 to 8 is crisp; we write $H = \{r \in \Re \mid 6 \le r \le 8\}$. Equivalently, H is described by its *membership* (or characteristic, or indicator) *function* (MF), $m_H : \Re \to \{0,1\}$, defined as

$$m_H(r) = \begin{Bmatrix} 1 & 6 \le r \le 8 \\ 0 & \text{otherwise} \end{Bmatrix}$$

The crisp set H and the graph of m_H are shown in the left half of Fig. 1. Every real number (r) either is in H or is not. Since m_H maps all real numbers $r \in \Re$ onto the two points (0,1), crisp sets correspond to two-valued logic: is or isn't, on or off, black or white, 1 or 0. In logic, values of m_H are called truth values with reference to the question, "Is r in H?" The answer is yes if and only if $m_H(r) = 1$; otherwise, no.

Consider next the set F of real numbers that are *close to 7*. Since the property "close to 7" is fuzzy, there is *not a unique* membership function for F. Rather, the modeler must decide, based on the potential application and properties desired for F, what m_F should be. Properties that might seem plausible for this F include (i) normality (MF(7) = 1), (ii) monotonicity (the closer r is to 7, the closer $m_F(r)$ is to 1, and conversely), and (iii) symmetry (numbers equally far left and right of 7 should have equal memberships). Given these intuitive constraints, either of the functions shown in the right half of Fig. 1 might be a useful representation of F. m_{F1} is discrete (the staircase graph), while m_{F2} is continuous but not smooth (the triangle graph). One can easily construct a MF for F so that every number has some positive membership in F, but we would not expect numbers "far from 7," 20 000 987 for example, to have much! One of the biggest differences between crisp and fuzzy sets is that the former always have unique MFs, whereas every fuzzy set has an infinite number of MFs that may represent it. This is at once both a weakness and a strength; uniqueness is sacrificed, but this gives a concomitant gain in terms of flexibility, enabling fuzzy models to be "adjusted" for maximum utility in a given situation.

In conventional set theory, sets of real objects, such as the numbers in H, are equivalent to, and isomorphically described by, a unique membership function such as m_H. However, there is no set-theory equivalent of "real objects" corresponding to m_F. Fuzzy sets are always (and only) *functions*, from a "universe of objects," say X, into [0,1]. This is depicted in Fig. 2, which illustrates that the fuzzy set is the *function* m_F that carries X into [0,1]. As defined, *every* function $m : X \to [0,1]$ is a fuzzy set. While this is true in a formal mathematical sense, many functions that qualify on this ground cannot be suitably interpreted as realizations of a conceptual fuzzy set. In other words, functions that map X into the unit interval *may* be fuzzy sets, but *become* fuzzy sets when, and only when, they match some intuitively plausible semantic description of imprecise properties of the objects in X.

One of the first questions asked about this scheme, and the one that is still asked most often, concerns the **relationship of fuzziness to probability**. Are fuzzy sets just a clever

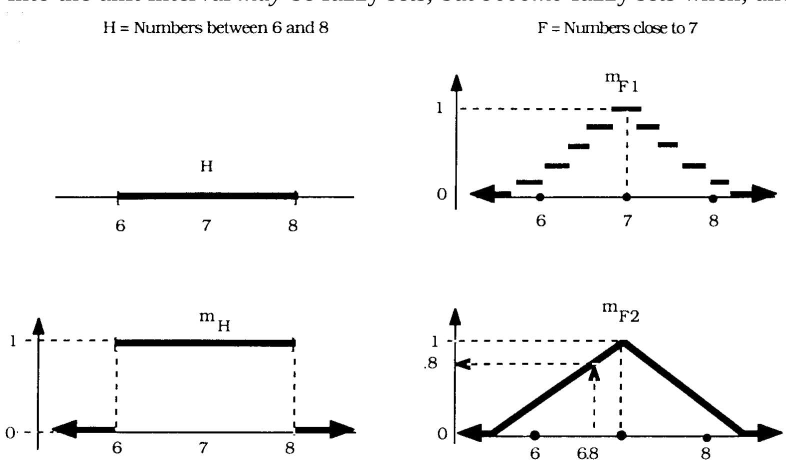

Fig. 1. Membership functions for hard and fuzzy subsets of $\Re$.

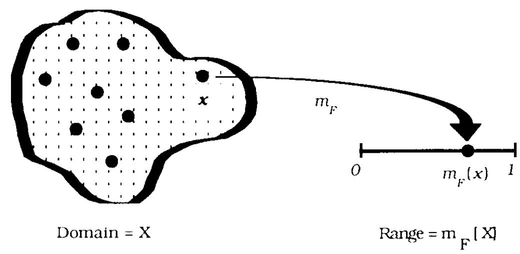

Fig. 2. Fuzzy sets are membership functions.

disguise for statistical models? Well, in a word, NO. Perhaps an example will help.

Example 1: Let the set of all liquids be the universe of objects, and let fuzzy subset L = {all *potable* (="suitable for drinking") liquids}. Suppose you had been in the desert for a week without drink and you came upon two bottles, A and B. You are told that the (fuzzy) membership of the liquid in A to L is 0.9 and also that the probability that the liquid in B belongs to L is 0.9. In other words, A contains a liquid that is potable with *degree of membership* 0.9, while B contains a liquid that is potable with *probability* 0.9. Confronted with this pair of bottles and given that you must drink from the one that you choose, which would *you* choose to drink from first? Why? Moreover, after an observation is made regarding the content of both bottles what are the (possible) values for membership and probability? [The answers to this "riddle" will be discussed in class].

Another common misunderstanding about fuzzy models over the years has been that they were offered as *replacements* for crisp (or probabilistic) models. To expand on this, first note from Figs. 1 and 2 that every crisp set is fuzzy, but not conversely. Most schemes that use the idea of fuzziness use it in this sense of *embedding*; that is, we work at preserving the conventional structure, and letting it dominate the output whenever it can, or whenever it must. Another example will illustrate this idea.

Example 2: Consider the plight of early mathematicians, who knew that the Taylor series for the real (bell-shaped) function $f(x) = 1 / (1 + x^2)$ was divergent at $x = \pm 1$ but could not understand why, especially since f is differentiable infinitely often at these two points. As is common knowledge for any student of complex variables nowadays, the *complex* function $f(z) = 1 / (1 + z^2)$ has poles at $z = \pm i$, two purely imaginary numbers. Thus, the complex function, which is an embedding of its real antecedent, cannot have a convergent power series expansion anywhere on the boundary of the unit disk in the plane; in particular at $z = \pm 0i \pm 1$, i.e., at the real numbers $x = \pm 1$. This exemplifies a general principle in mathematical modeling: given a real (seemingly insoluble) problem; enlarge the space, and look for a solution in some "imaginary" superset of the real problem; finally, specialize the "imaginary" solution to the original real constraints.

In Example 2 we spoke of "complexifying" the function f by embedding the real numbers in the complex plane, followed by "decomplexification" of the more general result to solve the original problem. Most fuzzy models follow a very similar pattern. Real problems that exhibit non-statistical uncertainty are first "fuzzified," some type of analysis is done on the larger problem, and then the results are specialized back to the original problem. In Example 2 we might call the return to the real line decomplexifying the function; in fuzzy models, this part of the procedure has come to be known as defuzzification. Defuzzification is usually necessary, of course, because even though we instruct a student to "apply the brakes pretty soon," in fact, the brake pedal must be operated crisply, at some real time. In other words, we cannot admonish a motor to "speed up a little," even if this instruction comes from a fuzzy controller we must alter its voltage by a specific amount. Thus defuzzification is both natural and necessary. Example 2 illustrates that this is hardly an idea that is novel; instead, we should regard it as a device that is useful.

Example 3: As a last, and perhaps more concrete, example about the use of fuzzy models, consider the system shown in Fig. 3, which depicts a simple inverted pendulum free to rotate in the plane of the figure on a pivot attached to the cart. The control problem is to keep the pendulum vertical at all times by applying a restoring force (control signal) $F(t)$ to the cart at some discrete times (t) in response to changes in both the linear and angular position and velocity of the pendulum. This problem can be formulated many ways. In one of the simpler versions used in conventional control theory, linearization of the equations of motion results in a model of the system whose stability characteristics are determined by examination of the *real parts* of the eigenvalues $\{\lambda_i\}$ of a 4×4 matrix of system constants. The lower track in Fig. 3 represents this case. It is well known that the pendulum can be stabilized by requiring $\mathrm{Re}(\lambda_i) < 0$, as shown in the middle of the lower track. This procedure is so commonplace in control engineering that most designers don't even think about the use of imaginary numbers to solve real problems, but it is clear that this process is exactly the same as was illustrated in Example 2 – a real problem is solved by temporarily passing to a larger, imaginary setting, analyzing the situation in the superset, and then specializing the result to get the desired answer.

The upper track in Fig. 3 depicts an alternative solution to this control problem that is based on fuzzy sets. This approach to stabilization of the pendulum is also well known, and yields a solution that in some ways is much better; e.g., the fuzzy controller is much less sensitive to changes in parameters such as the length and mass of the pendulum [2]. Note again the embedding principle: fuzzify, solve, defuzzify, control. The point of Example 3? Fuzzy models aren't really that different from more familiar ones. Sometimes they work better, and

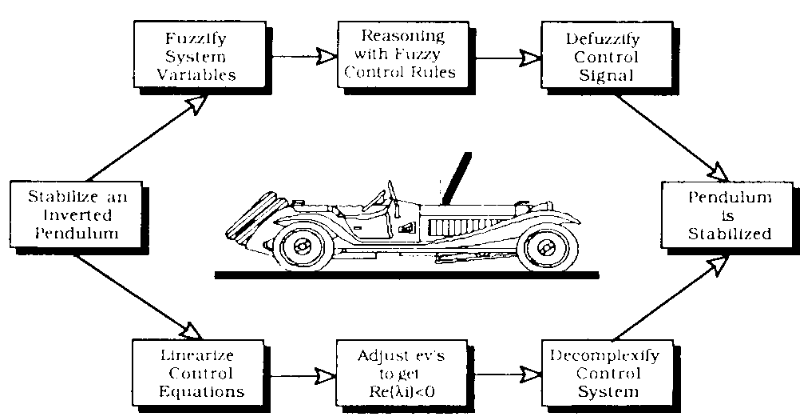

Fig. 3. Conventional and fuzzy solutions to real control problems found by embedding them in "imaginary" supersets.

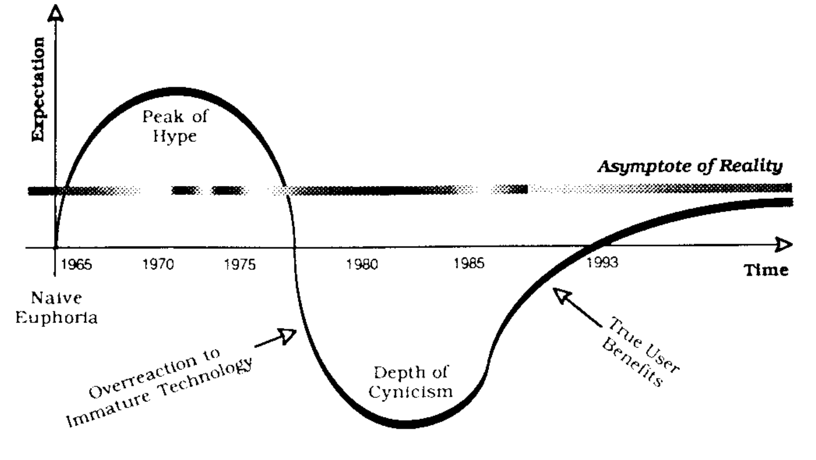

Fig. 4. Evolution of new technologies—time axis specialized to fuzzy models.

sometimes not. This is really the only criterion that should be used to judge any model, and there is much evidence nowadays that fuzzy approaches to real problems are often a good alternative to more familiar schemes. This is the point to which our discussion now turns.

Lets now discuss a little bit about the history of fuzzy sets. The enormous success of commercial applications which are at least partially dependent on fuzzy technologies fielded (in the main) by Japanese companies has led to a surge of curiosity about the utility of fuzzy logic for scientific and engineering applications. Over the last five or ten years, fuzzy models have supplanted more conventional technologies in many scientific applications and engineering systems, especially in control systems and pattern recognition. A recent *Newsweek* article indicates that the Japanese now hold thousands of patents on fuzzy devices used in applications as diverse as washing machines, TV camcorders, air conditioners, palm-top computers, vacuum cleaners, ship navigators, subway train controllers, and automobile transmissions [3]. It is this wealth of deployed, successful applications of fuzzy technology that is, in the main, responsible for current interest in the subject area.

Since 1965, many authors have generalized various parts of subdisciplines in mathematics, science, and engineering to include fuzzy cases. However, interest in fuzzy models was not really very widespread until their utility in fielded applications became apparent. The reasons for this delay in interest are many, but perhaps the most accurate explanation lies with the salient facts underlying the development of any new technology, which is succintly captured in Fig. 4. The horizontal axis of Fig. 4 is time, and the vertical axis is expectation – whose expectation? Well, usually, of the people who pay for development of the technology; but here I encourage you to interpret this axis in a much broader sense, for utility is, of course, in the eye of the user. The crucial part of Fig. 4 is the *asymptote of reality*, which bounds the delivery of the technology to a much lower expected value than early users project for it. The years shown along the time axis pertain to fuzzy models, and are, of course, approximate at best (with the exception of the initial one). When you look at this figure, you may enjoy deleting these years, and substituting your favorite new technology for the one illustrated. Each technology has its own evolution, and not all of them follow the pattern suggested by Fig. 4 (but you may be surprised to see how many do!). For example, try putting dates and identifying the people and events associated with, say, computational neural networks (which has an atypical, bimodal graph!); or artificial intelligence; or fractals; or complex numbers; and so on. Every new technology begins with naive euphoria -- its inventor(s) are usually submersed in the ideas themselves; it is their immediate colleagues that experience most of the wild enthusiasm. Most technologies are over promised, more often than not simply to generate funds to continue the work, for funding is an integral part of scientific development; without it, only the most imaginative and revolutionary ideas make it beyond the embryonic stage. Hype is a natural handmaiden to over promise, and most technologies build rapidly to a peak of hype. Following this, there is almost always an overreaction to ideas that are not fully developed, and this inevitably leads to a crash of sorts, followed by a period of wallowing in the depths of cynicism. Many new technologies evolve to this point, and then fade away. The ones that survive do so because someone finds a good use (= true user benefit) for the basic ideas. What constitutes a "good use"? For example, there are now many "good uses" in real systems for the complex numbers, as we have seen in Examples 2 and 3, but not many mathematicians thought so when Wessel, Argand, Hamilton, and Gauss made imaginary numbers sensible from a geometric point of view in the later 1800s. And in the context of fuzzy models, of course, "good use" corresponds to the plethora of products alluded to above. Interest in fuzzy systems in academia, industry, and government is also manifested by the rapid growth of national and international conferences. As noted above, successful applications of fuzzy models have gained great visibility through commercial applications in Japan. MITI in Japan started LIFE (Laboratory of Industrial Fuzzy Engineering) in 1988 with an annual budget of about $ 24 000 000 (U.S. dollars) for seven years. […]

Fuzzy Sets Theory

(Edited from J.-S.R. Jang, C.-T. Sun, and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Ch. 2, Prentice Hall, 1997)

Let X be a space of objects and x be a generic element of X. A classical set A, $A \subseteq X$, is defined as a collection of elements or objects $x \in X$, such that each element (x) can either belong or not to the set A. By defining a **characteristic** (or **membership**) **function** for each element x in X, we can represent a classical set A by a set of ordered pairs $(x,0)$ or

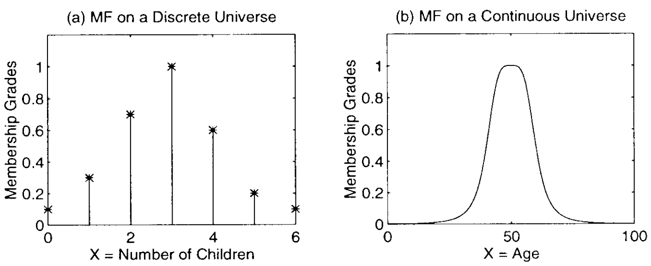

Figure 5. (a) B = "desirable number of children in a family"; (b) C = "about 50 years old"

(x,1), which indicates $x \notin A$ or $x \in A$, respectively. Unlike the aforementioned conventional set, a fuzzy set expresses the degree to which an element belongs to a set. Hence the membership function of a fuzzy set is allowed to have values between 0 and 1, which denote the degree of membership of an element in the given set.

Definition 1. *Fuzzy sets and membership functions*. If X is a collection of objects denoted generically by x, then a **fuzzy set** A in X is defined as a set of ordered pairs $A = \{(x,\mu_A(x)) \mid x \in X\}$, where, $\mu_A(x)$ is called the **membership function** (or MF for short) for the fuzzy set A. The MF maps each element of X to a membership grade (or membership value) <u>between</u> 0 and 1 (included).

Obviously, the definition of a fuzzy set is a simple extension of the definition of a classical (crisp) set in which the characteristic function is permitted to have any values between 0 and 1. If the value of the membership function is restricted to either 0 or 1, then A is reduced to a classical set. For clarity, we shall also refer to classical sets as ordinary sets, crisp sets, non-fuzzy sets, or just sets. Usually X is referred to as the **universe of discourse,** or simply the **universe,** and it may consist of discrete (ordered or non-ordered) objects or it can be a continuous space. This can be clarified by the following examples.

Example 1. *Fuzzy sets with a discrete non-ordered universe*. Let X = {San Francisco, Boston, Los Angeles} be the set of cities one may choose to live in. The fuzzy set A = "desirable city to live in" may be described as follows: A = {(San Francisco, 0.9), (Boston, 0.8), (Los Angeles, 0.6)}. Apparently the universe of discourse X is discrete and it contains non-ordered objects – in this case, three big cities in the United States. As one can see, the foregoing membership grades listed above are quite subjective; anyone can come up with three different but legitimate values to reflect his or her preference.

Example 2. *Fuzzy sets with a discrete ordered universe*. Let $X = \{0, 1, 2, 3, 4, 5, 6\}$ be the set of numbers of children a family may choose to have. Then the fuzzy set B = "desirable number of children in a family" may be described as follows: $B = \{(0, 0.1), (1, 0.3), (2, 0.7), (3, 1), (4, 0.7), (5, 0.3), (6, 0.1)\}$. Here we have a discrete ordered universe X; the MF for the fuzzy set B is shown in Fig. 5(a). Again, the membership grades of this fuzzy set are obviously subjective measures.

Example 3. *Fuzzy sets with a continuous universe*. Let $X = \Re^+$ be the set of possible ages for human beings. Then the fuzzy set C = "about 50 years old" may be expressed as $C = \{(x,\mu_C(x)) \mid x \in X\}$, where

$$\mu_C(x) = \frac{1}{1+\left(\dfrac{x-50}{10}\right)^4}$$

This is illustrated in Figure 5(b). From the preceding examples, it is obvious that the construction of a fuzzy set depends on two things: the identification of a suitable universe of discourse and the specification of an appropriate membership function. The specification of membership functions is subjective, which means that the membership functions specified for the same concept by different persons may vary considerably. This subjectivity comes from individual differences in perceiving or expressing abstract concepts and has little to do with randomness. Therefore, the **subjectivity** and **non-randomness** of fuzzy sets is the, primary difference between the study of fuzzy sets and probability theory, which deals with objective treatment of random phenomena.

In practice, when the universe of discourse X is a continuous space, we usually partition it into several fuzzy sets whose MFs cover X in a more or less uniform manner. These fuzzy sets, which usually carry names that conform to adjectives appearing in our daily linguistic usage, such as "large," "medium," or "small," are called *linguistic values* or *linguistic labels*. Thus, the universe of discourse X is often called the *linguistic variable*. An example on this follows.

Example 4. *Linguistic variables and linguistic values*. Suppose that X = "age." Then we can define fuzzy sets "young," "middle aged," and "old" that are characterized by MFs. Just as a variable can assume various values, a linguistic variable "age" can assume different linguistic values, such as "young," "middle aged," and "old" in this case. If "age" assumes the value of "young," then we have the expression "age is young," and so forth for the other values.

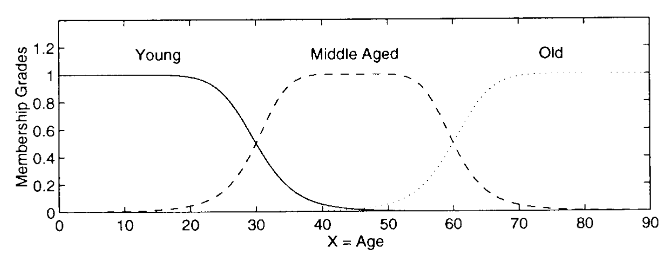

Figure 6. *Typical MFs of linguistic values "young," "middle aged," and "old."*

An example of MFs for these linguistic values are displayed in Fig. 6, where the universe of discourse X is totally covered by the MFs and the transition from one MF to another is smooth and gradual. Lets now define some nomenclature used in the literature.

Definition 2. *Support*. The **support** of a fuzzy set A is the set of all points x in X such that $\mu_A(x) > 0$.
Definition 3. *Core*. The **core** of a fuzzy set A is the set of all points x in X such that $\mu_A(x) = 1$.
Definition 4. *Normality*. A fuzzy set A is **normal** if its core is nonempty. In other words, we can always find at least a point $x \in X$ such that $\mu_A(x) = 1$.
Definition 5. *Crossover points*. A **crossover point** of a fuzzy set A is a point $x \in X$ at which $\mu_A(x) = 0.5$.
Definition 6. *Fuzzy singleton*. A fuzzy set whose support is a single point in X with $\mu_A(x) = 1$ is called a **fuzzy singleton.**
Definition 7. *α-cut, strong α-cut*. The **α-cut** or **α-level set** of a fuzzy set A is a crisp set defined by $A_\alpha = \{x \mid \mu_A(x) \geq \alpha\}$. **Strong α-cut** or **strong α-level set** are defined similarly: $A'_\alpha = \{x \mid \mu_A(x) > \alpha\}$.
Using this notation, we can express the support and core of a fuzzy set A as support(A) = A'_0 and core(A) = A_1.
Definition 8. *Convexity*. A fuzzy set A is **convex** if and only if for any $x_1, x_2 \in X$ and any $\lambda \in [0,1]$, $\mu_A(\lambda x_1 + (1-\lambda)x_2) \geq \min\{\mu_A(x_1), \mu_A(x_2)\}$. Alternatively, A is convex if all its α-level sets axe convex. Note that the definition of convexity of a fuzzy set is not as strict as the common definition of convexity of a function.
Definition 9. *Fuzzy numbers*. A fuzzy number A is a fuzzy set in the real line that satisfies the conditions for normality and convexity. Most fuzzy sets used in the literature satisfy the conditions for normality and convexity, so fuzzy numbers axe the most basic type of fuzzy sets.

Union, intersection, and complement are the most basic operations on classical sets. On the basis of these three operations, a number of identities can be established. Corresponding to the ordinary set operations of union, intersection, and complement, fuzzy sets have similar operations, which were initially defined in Zadeh's seminal paper [1]. Before introducing these three fuzzy set operations, first we shall define the notion of containment, which plays a central role in both ordinary and fuzzy sets. This definition of containment is, of course, a natural extension of the case for ordinary sets.
Definition 10. *Containment or subset*. Fuzzy set A is **contained** in fuzzy set B (or, equivalently, A is a **subset** of B, or A is smaller than or equal to B, $A \subseteq B$) if and only if $\mu_A(x) \leq \mu_B(x)$ for all x.
Definition 11. *Union (disjunction).* The **union** of two fuzzy sets A and B is a fuzzy set C, written as $C = A \cup B$ or $C = A$ OR B, whose MF is related to those of A and B by $\mu_C(x) = \max(\mu_A(x), \mu_B(x))$.
Definition 12. *Intersection (conjunction).* The **intersection** of two fuzzy sets A and B is a fuzzy set C, written as $C = A \cap B$ or $C = A$ AND B, whose MF is related to those of A and B by $\mu_C(x) = \min(\mu_A(x), \mu_B(x))$.
Definition 13. *Complement (negation).* The complement of fuzzy set A, denoted by $\bar{A}$ or NOT A, is defined as $\mu_{\bar{A}}(x) = 1 - \mu_A(x)$.

Note that the operations introduced in the last three definitions (11 to 13) perform exactly as the corresponding operations for ordinary sets if the values of the membership functions are restricted to either 0 or 1. However, it is understood that these functions are not the only possible generalizations of the crisp set operations. For each of the aforementioned three set operations, several different classes of functions with desirable properties have been proposed subsequently in the literature (e.g. algebraic sum for union and product for intersection). In general, union and intersection of two fuzzy sets can be defined through T-conorm (or S-norm) and T-norm operators respectively. These two operators are functions $S,T : [0,1]\times[0,1] \rightarrow [0,1]$ satisfying some convenient boundary, monotonicity, commutativity and associativity properties. As pointed out by Zadeh [1], a more intuitive but equivalent definition of union is the, “smallest” fuzzy set containing both A and B. Alternatively, if D is any fuzzy set that contains both A and B, then it also contains $A \cup B$. Analogously, the intersection of A and B is the "largest" fuzzy set which is contained in both A and B.

(Following is edited from J.M. Mendel, "Fuzzy Logic Systems for Engineering: A Tutorial," *Proc. of IEEE*, 83(**3**), 1995.)

The two fundamental (Aristotelian) laws of crisp set theory are: 1) *Law of Contradiction*: $A \cup \bar{A} = X$ (i.e., a set and its complement must comprise the universe of discourse), and 2) *Law of Excluded Middle*: $A \cap \bar{A} = \varnothing$ (i.e., an object can either be in its set or its complement; it cannot simultaneously be in both). It can be easily seen that for every fuzzy set that is non-crisp (i.e., whose membership function does not <u>only</u> assume values 0 and 1) both laws are broken (i.e., for fuzzy sets $A \cup \bar{A} \neq X$ and $A \cap \bar{A} \neq \varnothing$). Indeed $\forall x \in A$ such that $\mu_A(x) = \alpha$, $0 < \alpha < 1$: $\mu_{A \cup \bar{A}}(x) = \max\{\alpha, 1-\alpha\} \neq 1$ and $\mu_{A \cap \bar{A}}(x) = \min\{\alpha, 1-\alpha\} \neq 0$. For example, we can see from Fig. 6 how a 30 year old person is young with degree of membership 0.5 and not young with degree of membership 0.5. In fact, one of the ways to describe the difference between crisp set theory and fuzzy set theory is to explain that these two laws do not hold in fuzzy set theory. Consequently, any other mathematics that relies on crisp set theory, such as (frequency based) probability, must be different from fuzzy set theory.

We will now introduce the concept of *relations* in both crisp and fuzzy sets; this will later help us in approaching fuzzy logic. A **crisp relation** represents the *presence or absence* of association, interaction or interconnectedness between the elements of two or more sets. Given two sets X and Y a relation R between X and Y is itself a set $R(X,Y)$ subset of $X \times$

Y. For example the ordering relation "less than" (<) is a relation in $\Re^2$ defined as $LT(\Re,\Re) = \{(x,y) \mid x < y\}$. The point (1,123) belongs to $LT(\Re,\Re)$ while obviously (123,1) does not. (Note: by $X \times Y$ we indicate the Cartesian product of sets X and Y, that is the set of ordered couples with values from X and Y respectively, i.e., $X \times Y = \{(x,y) \mid x \in X \text{ and } y \in Y\}$.)

Definition 14. *Fuzzy relation.* A **fuzzy relation** represents a *degree of presence or absence* of association, interaction or interconnectedness between the elements of two or more sets. Some examples of (binary) fuzzy relations are: x is much larger than y, y is very close to x, z is much greener than y. Let X and Y be two universes of discourse. A fuzzy relation $R(X,Y)$ is a fuzzy set in the product space $X \times Y$, i.e., it is a fuzzy subset of $X \times Y$, and is characterized by the membership function $\mu_R(x,y)$, i.e., $R(X,Y) = \{((x,y),\mu_R(x,y)) \mid (x,y) \in X \times Y\}$.

The difference between a fuzzy relation and a crisp relation is that for the former any membership value in [0,1] is allowed while for the latter only 0 and 1 memberships are. This is why a fuzzy relation is expressing not only the interconnection between the elements of two or more sets (e.g., as a crisp relation does) but also the degree or extent of this association. Because fuzzy relations are fuzzy sets in product space, set theoretic operations can be defined for them using definitions 11 through 13.

Next, we consider the composition of crisp relations from different product spaces that share a common set, namely $P(X,Y)$ and $Q(Y,Z)$. The *composition* of these two relations is denoted by $R(X,Z) = P(X,Y) \circ Q(Y,Z)$ and is defined as a subset $R(X,Z)$ of $X \times Z$ such that $(x,z) \in R(X,Z)$ if and only if there exists at least one $y \in Y$ such that $(x,y) \in P(X,Y)$ and $(y,z) \in Q(Y,Z)$. This can be expressed in terms of characteristic functions through either the max-min or the max-product compositions respectively defined as

$$\mu_{P\circ Q}(x,z) = \max_{y}\left(\min\{\mu_P(x,y),\mu_Q(y,z)\}\right) \qquad \mu_{P\circ Q}(x,z) = \max_{y}\left(\mu_P(x,y)\mu_Q(y,z)\right)$$

The composition of fuzzy relations from different product spaces that share a common set is defined analogously to the crisp composition, except that in the fuzzy case the sets are fuzzy.

Definition 15. *Composition of fuzzy relations*. Given two relations $P(X,Y)$ and $Q(Y,Z)$ and their associated membership functions $\mu_P(x,y)$ and $\mu_Q(y,z)$, the **composition** of these two relations is denoted by $R(X,Z) = P(X,Y) \circ Q(Y,Z)$ (or simply R

$$\mu_{P\circ Q}(x,z) = \sup_{y\in Y}\left[\mu_P(x,y) \otimes \mu_Q(y,z)\right]$$

$= P \circ Q$) and is defined as a subset $R(X,Z)$ of $X \times Z$ defined by the membership function

Motivation for using the T-norm operator (⊗) comes from the crisp max-min and max-product compositions, because both the min and the product are T-norms. This is also sometimes referred to as sup-star composition due to an alternative symbol for T-norm (e.g., ★). Although it is permissible to use other T-norms, the most commonly used sup-star compositions are the sup-min and sup-product. Observe that, unlike the case of crisp compositions, for which exactly the same results are obtained using either the max-min or the max-product composition, the same results are not obtained in the case of fuzzy compositions. This is a major difference between fuzzy composition and crisp composition.

Suppose fuzzy relation P is just a fuzzy set, so that $\mu_P(x,y)$ just becomes $\mu_P(x)$, e.g., "x is medium large and z is smaller than y." Then $Y = X$ and the membership function for the composition of P and Q becomes

$$\mu_{P\circ Q}(z) = \sup_{x\in X}\left[\mu_P(x) \otimes \mu_Q(x,z)\right]$$

Note that now this is only a function of the variable z. This equation will be useful in later developments of fuzzy reasoning.

Fuzzy Logic

It is well established that prepositional logic is isomorphic to set theory under the appropriate correspondence between components of these two mathematical systems. Furthermore, both of these systems are isomorphic to a Boolean algebra, which is a mathematical system, defined by abstract entities and their axiomatic properties. The isomorphism between Boolean algebra, set theory, and propositional logic guarantees that every theorem in any one of these theories has a counterpart in each of the other two theories. These isomorphisms allow us, in effect, to cover all these theories by developing only one of them. Consequently, we will not spend a lot of time reviewing crisp logic; but we must spend some time on it, especially on the concept of implication, in order to reach the comparable concept in fuzzy logic.

Rules are a form of propositions. A *proposition* is an ordinary statement involving terms which have been defined, e.g., "The damping ratio is low." Consequently, we could have the following rule: "IF the damping ratio is low, THEN the system's impulse response oscillates a long time before it dies out." In traditional propositional logic, a proposition must be meaningful to call it "true" or "false," whether or not we know which of these terms properly applies. Logical reasoning is the process of combining given propositions into other propositions, and then doing this over and over again. Propositions can be combined in many ways, all of which are derived from three fundamental operations: *conjunction* (denoted $p \wedge q$), where we assert the simultaneous truth of two separate propositions p and q; *disjunction* ($p \vee q$), where we

Table 1. Truth table for five operations that are frequently applied to propositions

p	q	$p \wedge q$	$p \vee q$	$p \rightarrow q$	$p \leftrightarrow q$	$\sim p$
T	T	T	T	T	T	F
T	F	F	T	F	F	F
F	T	F	T	T	F	T
F	F	F	F	T	T	T

assert the truth of either or both of two separate propositions; and *implication* ($p \rightarrow q$) which usually takes the form of an IF-THEN rule. The IF part of an implication is called the *antecedent*, whereas the THEN part is called the *consequent*. In addition to generating propositions using conjunction, disjunction or implication, a new proposition can be obtained from a given one by prefixing the clause "it is false that …". This is the operation of *negation* ($\sim p$). Additionally, $p \leftrightarrow q$ is the *equivalence* relation; it means that p and q are both true or false.

In traditional propositional logic we combine unrelated propositions into an implication, and *we do not assume any cause or effect relation to exist*. We will see later that this last statement causes serious problems when we try to apply traditional propositional logic to engineering applications, where cause and effect rule (i.e., a (causal) system does not respond until an input is applied to it). In traditional propositional logic an implication is said to be true if one of the following holds (see also Table 1): 1) (antecedent is true, consequent is true), 2) (antecedent is false, consequent is false), 3) (antecedent is false, consequent is true); the implication is called false when 4) (antecedent is true, consequent is false). Situation 1) is the familiar one of common experience. Situation 2) is also reasonable, for if we start from a false assumption we expect to reach a false conclusion, however, intuition is not always reliable. We may reason correctly from a false antecedent to a true consequent (e.g., IF 1 = 2 is false, but, adding 2 = 1 to this false statement, lets us correctly conclude that 3 = 3); hence, a false antecedent can lead to a consequent which is either true or false, and thus both situations 2) and 3) are allowed in traditional propositional logic. Finally, situation 4) is in accord with our intuition, for an implication is clearly false if a true antecedent leads to a false consequent. A logical structure is constructed by applying the above four operations to propositions. The objective of a logical structure is to determine the truth or falsehood of all propositions which can be stated in the terminology of this structure. *A truth table* is very convenient for showing relationships between several propositions. The fundamental truth tables for conjunction, disjunction, implication, equivalence and negation are collected together in Table 1, in which symbol T means that the corresponding proposition is true, and symbol F that it is false. The fundamental axioms of traditional propositional logic are: 1) every proposition is either true or false, but not both true and false, 2) the expressions given by defined terms are propositions, and, 3) the truth table (in Table 1) for conjunction, disjunction, implication, equivalence, and negation. Using truth tables, we can derive many interpretations of the preceding operations and can also prove relationships about them.

A *tautology* is a proposition formed by combining other propositions, which is true regardless of the truth or falsehood of the forming propositions. The most important tautologies for our work are: $(p \rightarrow q) \leftrightarrow \sim[p \wedge (\sim q)] \leftrightarrow (\sim p) \vee q$. These tautologies can be verified by substituting all the possible combinations for p and q and verifying how the equivalence always holds true. The importance of these tautologies is that they let us express the membership function for $p \rightarrow q$ in terms of membership functions of either propositions p and $\sim q$ or $\sim p$ and q. Thus giving us

$$\mu_{p\rightarrow q}(x,y) = 1 - \mu_{p\cap\bar{q}}(x,y) = 1 - min\{\mu_p(x), 1-\mu_q(y)\}$$
$$\mu_{p\rightarrow q}(x,y) = \mu_{\bar{p}\cup q}(x,y) = max\{1-\mu_p(x), \mu_q(y)\}$$

Note that instead of min and max we could have used product and algebraic sum for intersection and union respectively. These two equations can be verified by substituting 1 for true and 0 for false.

In traditional propositional logic there are two very important inference rules, *Modus Ponens* and *Modus Tollens. Modus Ponens*: *Premise 1:* "x is A"; *Premise 2:* "IF x is A THEN y is B", *Consequence:* "y is B." Modus Ponens is associated with the implication "A implies B." In terms of propositions p and q, Modus Ponens is expressed as : $(p \wedge (p \rightarrow q)) \rightarrow q$. *Modus Tollens: Premise 1*: "y is not B"; *Premise 2*: "IF x is A THEN y is B"; *Consequence:* "x is not A." In terms of propositions p and q, Modus Tollens is expressed as $((\sim q) \wedge (p \rightarrow q)) \rightarrow (\sim p)$. Whereas Modus Ponens plays a central role in engineering applications of logic, due in large part to cause and effect, Modus Tollens does not seem to have yet played much of a role.

Fuzzy logic begins by borrowing notions from crisp logic, just as fuzzy set theory; however, as we shall demonstrate below, doing this is inadequate for engineering applications of fuzzy logic, because, in engineering, cause and effect is the cornerstone of modeling, whereas in traditional propositional logic it is not. Ultimately, this will cause us to define "engineering" implication operators. Before doing this, let us develop an understanding of why the traditional approach fails us in engineering. As in our extension of crisp set theory to fuzzy set theory, our extension of crisp logic to fuzzy logic is made by replacing the bivalent membership functions of crisp logic with fuzzy membership functions. That is all there is to it; hence, the IF-THEN statement "IF x is A, THEN y is B," where $x \in X$ and $y \in Y$, has a membership function $\mu_{A\rightarrow B}(x,y) \in [0,1]$. Note that $\mu_{A\rightarrow B}(x,y)$ measures the degree of truth of the implication relation between x and y. This membership function can be defined as for the crisp case above. In fuzzy logic, Modus Ponens is extended to *Generalized Modus Ponens*: *Premise 1:* "x is A^*"; *Premise 2:* "IF x is A THEN y is B"; *Consequence*: "y is B^*." Compare Modus

Ponens and Generalized Modus Ponens to see their subtle differences, namely, in the latter, fuzzy set A^* is not the necessarily the same as rule antecedent fuzzy set A, and fuzzy set B^* is not necessarily the same as rule consequent B.

Example 4: Consider the rule "IF a man is short, THEN he will not make a very good professional basketball player." Here fuzzy set A is *short man* and fuzzy set B is *not a very good professional basketball player.* We are now given Premise 1, as "This man is under five feet tall." Here A^* is the fuzzy set *man under five feet tall*. Clearly A and A^* are different but similar. We now draw the following consequence: "He will make a poor professional basketball player." Here B^* is the fuzzy set *poor professional basketball player*, and it is different from B, although they are indeed similar. Note how Premise 1 could have been "This man is five feet tall" (this would correspond to a fuzzy singleton) and we would have reached the same conclusion.

We see that in crisp logic a rule will be fired only if the first premise is exactly the same as the antecedent of the rule, and, the result of such rule firing is the rule's actual consequent. In fuzzy logic, on the other hand, a rule is fired so long as there is a nonzero degree of similarity between the first premise and the antecedent of the rule, and the result of such rule firing is a consequent that has nonzero degree of similarity to the rule' consequent.

Generalized Modus Ponens is a fuzzy composition where the first fuzzy relation is merely the fuzzy set, A^*. Consequently, $\mu_{B^*}(y)$ is obtained from the sup-star composition as

$$\mu_{B^*}(y) = \sup_{x \in A^*} \left[\mu_{A^*}(x) \otimes \mu_{A \to B}(x,y)\right]$$

Lets now think at an application of this approach. Given an observation x_1 we want to determine what is the correct action y_1 corresponding to the observation. This observation needs to correspond to the first premise in generalized modus ponens, thus it needs to be a fuzzy set (e.g., A^*). But it really is a crisp number, thus it needs to first be transformed into a fuzzy set (*fuzzification*) the rule (or rules) is then processed thus producing an output fuzzy set (B^*) that needs to be transformed into a crisp number (*defuzzification*) to be useful in the real world. The operations that we just described correspond to the mode of functioning of a fuzzy logic system (see also Fig. 3 and the related discussion). Thus in a FLS, an input is fuzzified, then processed by a rule base through an inference process and finally defuzzified to produce a usable (crisp) output. There are several types of fuzzifiers and defuzzifiers and their discussion is outside the scope of this introduction. One of the most popular types of fuzzifiers is the singleton fuzzifier. In this fuzzification scheme an observation x_1 is transformed into a fuzzy set being a singleton with support $\{x_1\}$. Thus, with the nomenclature introduced above $\mu_{A^*}(x)$ is zero everywhere besides at $x = x_1$. Consequently $\mu_{B^*}(y)$ becomes

$$\mu_{B^*}(y) = \sup_{x \in A^*} \left[\mu_{A^*}(x) \otimes \mu_{A \to B}(x,y)\right] = 1 \otimes \mu_{A \to B}(x_1, y) = \mu_{A \to B}(x_1, y) = 1 - \min\{\mu_A(x_1), 1 - \mu_B(y)\}$$

A graphical interpretation of this equation using a triangular membership function for $\mu_B(y)$ (a common choice in FLSs) reveals a disturbing result for an engineering application. It tells us that, given a specific input $x = x_1$, the result of firing a specific rule, whose consequent is associated with a specific fuzzy set of finite support (the base of the triangle), is a fuzzy set whose support is infinite. (Please try this construction yourself, we can also discuss it further in class.) Clearly, this does not make much sense from an engineering perspective, where cause (e.g., system input) shoud lead to effect (e.g., system output), and noncause should not lead to anything. Any other choice for $\mu_{A \to B}(x,y)$ (i.e., among the ones illustrated above) leads us to similar results. Mamdani [4] seems to have been the first one to recognize the problem we have just demonstrated, although he does not explain it the way we have. He chose to work with a *minimum implication* defined as $\mu_{A \to B}(x,y) = \min\{\mu_A(x), \mu_B(y)\}$. His reason for choosing this definition do not seem to be based on cause and effect, but, instead on simplicity of computation. Later, Larsen [5] proposed a *product implication* defined as $\mu_{A \to B}(x,y) = \mu_A(x)\ \mu_B(y)$. Again, the reason for this choice was simplicity of computation rather than cause and effect. Today, minimum and product inferences are the most widely used inferences in the engineering applications of fuzzy logic; but, what do they have to do with traditional propositional logic? It can be easily seen that neither minimum inference nor product inference agree with the accepted propositional logic definition of implication, given in Table 1. Hence, minimum and product inferences have nothing to do with traditional propositional logic. Interestingly enough minimum and product inferences preserve cause and effect, i.e. the implication is fired only when the antecedent and the consequent are both true. Thus, they are sometimes collectively referred to as *engineering implications*.

A very interesting development of FLSs is that they can be regarded as a parametric nonlinear mapping that was proven to be a universal approximator. Thus, an ongoing field of research is that of learning methods for FLSs treated as black boxes [6]. The advantages offered by FLSs are twofold: 1) they can be better initialized from human experts with rules in linguistic form (as compared to other black box models); 2) the final result of the learning process is easily interpreted and sometimes can reveal some hidden system characteristics, thus maybe showing a hope for the very longed and real artificial intelligence where a computer system could actually teach something to a human. Even though FLSs are definitely the most popular outcome of fuzzy sets and fuzzy logic, a lot of other research and application fields sprung out of them. Among some of them there is fuzzy optimization and fuzzy linear programming, fuzzy preference modeling and fuzzy multi attribute decision making, linguistic modeling and decision models (a survey on fuzzy sets in OR can be found in [7]).

REFERENCES

[1] L.A. Zadeh, "Fuzzy Sets," *Information and Control*, Vol. 8, pp. 338-352, 1965.

[2] C.C. Lee, "Fuzzy Logic in Control Systems: Fuzzy Logic Controllers," (parts I and II), *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 20, No. 2, pp. 404-435, 1990.

[3] "The future is fuzzy," *Newsweek*, May 1990.

[4] E.H. Mamdani, "Applications of fuzzy algorithms for simple dynamic plant," *Proc. IEE*, 121, pp. 1585-1588, 1974.

[5] P.M. Larsen, "Industrial Applications of Fuzzy Logic Control," *International Journal of Man, machine Studies*, Vol. 12, No. 1, pp. 3-10, 1980.

[6] T. Takagi, and M. Sugeno, "Fuzzy identification of Systems and its Applications to Modeling and Control," *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 15, No. 1, January-February 1985.

[7] F. Herrera, and J.L. Verdegay, "Fuzzy Sets and Operations Research: Perspectives," *Fuzzy Sets and Systems*, Vol. 90, pp. 207-218, 1997.