On Measures of Fuzziness

J. Knopfmacher

University of the Witwatersrand, Johannesburg, South Africa

Submitted by L. A. Zadeh

1. INTRODUCTION

Work of L. A. Zadeh and subsequent authors (see e.g., [1–3, 5–7] and their references) has established the wide-spread significance of the concept of a "fuzzy" set, particularly in applications of mathematics. By definition, a fuzzy set in a set $X$ is any mapping $f: X \rightarrow L$, where $L$ is a given lattice with least element 0 and greatest element $1 \neq 0$. The value $f(x)$ [$x \in X$] assigns to $x$ its "grade of membership" in the fuzzy set $f$; an ordinary set $A \subseteq X$ is specified by its characteristic function $\chi_A: X \rightarrow \{0, 1\}$ and, by contrast with with the general case, such a function $\chi_A$ may be called a sharp set in $X$. (For more detailed motivation, see the references indicated above.)

De Luca and Termini [2] raised the interesting question of assigning to any fuzzy set $f$ in $X$ some measure of its "fuzziness," the degree $d(f)$ of fuzziness of $f$. For a quantitative measure, it is reasonable firstly to try to define a real number $d(f)$ with suitable properties, and to restrict attention to the original functions considered by Zadeh, for which $L$ is the ordinary real unit interval $I = [0, 1]$. (In some situations, one might wish to assign independent or incomparable grades of membership to the elements of $X$—e.g., weight and color. Such cases could often easily be reduced to the present one, by considering a cartesian product $L = I^n$ of $I$ and a vector-valued measure $d(f)$.)

In addition to the above, it is not unreasonable to place some "quantitative" restrictions on $X$ or $f$, or both. Thus De Luca and Termini consider only a finite set $X$ (with $f: X \rightarrow I$ arbitrary), and they raise the question of extending their considerations to infinite sets $X$. This note considers the case when $X$ is an arbitrary set with a totally finite positive measure $\mu$ defined on a $\sigma$-algebra ($\mathcal{F}$) of subsets of $X$, and $f: X \rightarrow I$ is a measurable function. (All measure-theoretic terms and results used here may be found in Halmos's book [4].) The case of [2] is covered by letting $X$ be finite, $\mathcal{F}$ be the Boolean algebra of all subsets of $X$ and $\mu(x) = 1$ for $x \in X$; here all maps $f: X \rightarrow I$ are measurable.
In seeking to define \( d(f) \) when \( X \) is finite, De Luca and Termini lay down three intuitively reasonable properties that such a measure should have:

\( P_1: \ d(f) = 0 \) if and only if \( f \) is a sharp set in \( X \).

\( P_2: \ d(f) \) assumes a unique maximum value for fuzzy sets in \( X \) if and only if \( f \) is the constant function \( f_{1/2} \) taking value \( \frac{1}{2} \).

\( P_3: \ d(f) \geq d(f^*) \) whenever \( f^* \) is a "sharpened" version of \( f \), i.e., a fuzzy set such that \( f^*(x) > f(x) \) if \( f(x) > \frac{1}{2} \), and \( f^*(x) < f(x) \) if \( f(x) < \frac{1}{2} \).

By first introducing a function \( H(f) \) formally similar to the entropy of a finite probability distribution, De Luca and Termini define a function \( d(f) \) satisfying \( P_1 \) to \( P_3 \) and also:

\( P_4^*: \ d(f) = d(\tilde{f}) \) where \( \tilde{f} \) denotes the "complement" of \( f \), i.e., the function such that \( \tilde{f}(x) = 1 - f(x) \) \( \forall x \in X \).

\( P_5^*: \ d(f) \) is a nonnegative "valuation" on the lattice of all fuzzy sets, i.e.:

\[
d(f \lor g) + d(f \land g) = d(f) + d(g),
\]

where

\[
(f \lor g)(x) = \max(f(x), g(x)), \quad (f \land g)(x) = \min(f(x), g(x)).
\]

Here \( P_4^* \) is a very natural property to expect of \( d(f) \) while \( P_5^* \), if less intuitive, is nevertheless very useful.

It will be shown below that there are infinitely many quite different degree functions \( d(f) \) that have all the above properties, and this is so even when \( f \) is any measurable fuzzy set in an arbitrary measure space \((X, \mathcal{S}, \mu)\) with \( 0 < \mu(X) < \infty \) provided that \( P_1 \) and \( P_2 \) are modified slightly as below. In addition, some further intuitively reasonable stipulations for \( d(f) \) are noted, and it is shown that the various properties considered characterize \( d(f) \) uniquely up to membership of a certain explicitly defined class of functions.

2. Additional Properties of \( d(f) \)

Now let \((X, \mathcal{S}, \mu)\) denote an arbitrary measurable space, with \( \mathcal{S} \) a \( \sigma \)-algebra and \( 0 < \mu(X) < \infty \), and let attention be restricted to the set \( \mathcal{F}(X) \) of all fuzzy sets \( f: X \rightarrow I \) that are measurable as real-valued functions. In this general setting, it appears necessary to rewrite \( P_1 \) and \( P_2 \) slightly. The following statements reduce to \( P_1 \) and \( P_2 \) in the finite case considered previously.

\( P_1^*: \ d(f) = 0 \) if and only if \( f \) is sharp almost everywhere (a.e.).

\( P_2^*: \ d(f) \) assumes a unique maximum value for measurable fuzzy sets in \( X \) if and only if \( f \) coincides a.e. with the constant function \( f_{1/2} \) taking value \( \frac{1}{2} \).
Next, a little reflection on $P_3$ suggests that its conclusion should be valid under slightly weaker assumptions:

$P_3^*$: $d(f) \geq d(f^*)$ whenever $f^* \in \mathcal{F}(X)$ has the property that $f^*(x) \geq f(x)$ if $f(x) > \frac{1}{2}$, and $f^*(x) \leq f(x)$ if $f(x) < \frac{1}{2}$.

Intuitively, as with the entropy of a probability distribution, one feels that a slight variation in the values of a fuzzy set $f$ should have little effect on $d(f)$. In order to make this precise, consider the uniform metric $\rho$ on $\mathcal{F}(X)$ defined by

$$\rho(f, g) = \sup_{x \in X} |f(x) - g(x)|.$$  

(In the earlier finite case, $\mathcal{F}(X)$ may be identified with $\mathbb{R}^N$ where $N = \text{card} X$, and here $\rho$ is one of the standard Euclidean-topology metrics restricted to $\mathbb{R}^N$.) We stipulate

$P_6^*$: $d(f)$ is a continuous function of $f \in \mathcal{F}(X)$ relative to the metric $\rho$.

It may be noted that, in the finite case, $P_6^*$ and the compactness of $\mathbb{R}^N$ imply that $d(f)$ must take a maximum value; $P_2^*$ implies that there is such a value in general, and that it is taken by essentially only one function $f$.

Now consider the constant function $f_\alpha$ taking value $\alpha \in I$. By $P_6^*$ and $P_2^*$, $\Delta(\alpha) = d(f_\alpha)$ is a continuous function of $\alpha$ taking the maximum value $\Delta(\frac{1}{2})$ at the single point $\alpha = \frac{1}{2}$. Thus $f_\alpha$ reaches maximum "fuzziness" exactly when $\alpha = \frac{1}{2}$, and it is perhaps reasonable to suppose that $d(f_\alpha)$ increases as $\alpha \to \frac{1}{2}$, and decreases as $\alpha \to 0$ or 1. Since $P_4^*$ implies that $\Delta(\alpha) = \Delta(1 - \alpha)$, these considerations may be expressed as

$P_7^*$: $\Delta(\alpha) = d(f_\alpha)$ is a strictly increasing function of $\alpha$ for $\alpha \in [0, \frac{1}{2}]$.

Lastly consider any $f \in \mathcal{F}(X)$. One might expect $d(f)$ to be affected mainly by those values $f(x) = \alpha$ that occur "most commonly" for $x \in X$. Since $(X, \mathcal{S}, \mu)$ is virtually a probability space, this remark may be formalized by

$P_8^*$: $d(f) = 1/\mu(X) \int \Delta(f(x)) \, d\mu(x)$, the "expected" value of the composition $\Delta f$.

3. Existence Theorem

**Theorem.** Let $(X, \mathcal{S}, \mu)$ denote any measure space with $\mathcal{S}$ a $\sigma$-algebra and $0 < \mu(X) < \infty$, and let $\mathcal{F}(X)$ denote the set of all fuzzy sets $f$ in $X$ that are measurable as real-valued functions. Finally let $\Delta$ denote an arbitrary real-valued function of $\alpha \in I$ such that $\Delta(0) = \Delta(1) = 0$, $\Delta(\alpha) = \Delta(1 - \alpha)$, and $\Delta$ is strictly increasing for $\alpha \in [0, \frac{1}{2}]$. Then the definition

$$d(f) = \frac{1}{\mu(X)} \int \Delta(f(x)) \, d\mu(x),$$

yields a function of $f \in \mathcal{F}(X)$ which satisfies all the properties $P_1^* \text{ to } P_8^*$. 
Proof. Properties \( P_4^* \), \( P_7^* \) and \( P_8^* \) are immediate. Also, if \( g = f \) a.e.
then \( \Delta g = \Delta f \) a.e., and so \( d(g) = d(f) \). Now, if \( f \) is sharp then
\[
\begin{align*}
    d(f) &= \frac{1}{\mu(X)} \left\{ \int_{f^{-1}(0)} \mu(x) + \int_{f^{-1}(1)} \Delta(f(x)) \, d\mu(x) \right\} \\
    &= \frac{1}{\mu(X)} \left\{ \int_{f^{-1}(0)} 0 \, d\mu + \int_{f^{-1}(1)} 0 \, d\mu \right\} = 0.
\end{align*}
\]
Therefore \( d(g) = 0 \) if \( g \) is sharp a.e. On the other hand, if \( f \in \mathcal{F}(X) \) is an
function with \( d(f) = 0 \) then \( \Delta f = 0 \) a.e., and hence \( f(x) = 0 \) or \( 1 \) a.e.
Thus \( f \) is sharp a.e., and \( P_1^* \) follows.

Next, for any \( f \in \mathcal{F}(X) \), \( \Delta(f(x)) \leq \Delta(\frac{1}{2}) \) [\( x \in X \)], i.e., \( \Delta f \leq \Delta f_{1/2} \). Hence
\( d(f) \leq d(f_{1/2}) = \Delta(\frac{1}{2}) \). Conversely, if \( d(g) = \Delta(\frac{1}{2}) \) then \( \Delta g = \Delta f_{1/2} \) a.e., and
this implies that \( f = f_{1/2} \) a.e., yielding \( P_2^* \).

Now consider a function \( f^* \) as in \( P_3^* \). Since \( \Delta f^* \leq \Delta f \),
\[
\begin{align*}
    d(f^*) &= \frac{1}{\mu(X)} \left\{ \int_{f^{-1}[0,\frac{1}{2}]} + \int_{f^{-1}[\frac{1}{2},1]} \Delta(f^*(x)) \, d\mu(x) \right\} \\
    &\leq \frac{1}{\mu(X)} \int \Delta(f(x)) \, d\mu(x) = d(f).
\end{align*}
\]

For \( P_5^* \), note that \( f \vee g = f \) when \( f \geq g \), i.e., \( f-g \geq 0 \), and \( f \vee g = g \)
when \( f < g \), i.e., \( f-g < 0 \). A similar comment applies to \( f \wedge g \), and, since
\( h = f - g \) is a measurable real-valued function on \( X \), it follows that
\[
\begin{align*}
    d(f \vee g) + d(f \wedge g) &= \frac{1}{\mu(X)} \left\{ \int_{h^{-1}[0,\infty]} \Delta(f(x)) \, d\mu(x) + \int_{h^{-1}(-\infty,0]} \Delta(g(x)) \, d\mu(x) \right\} \\
    &\quad + \int_{h^{-1}[0,\infty)} \Delta(g(x)) \, d\mu(x) + \int_{h^{-1}(-\infty,0]} \Delta(f(x)) \, d\mu(x) \right\} \\
    &= \frac{1}{\mu(X)} \left\{ \int \Delta(f(x)) \, d\mu(x) + \int \Delta(g(x)) \, d\mu(x) \right\} \\
    &= d(f) + d(g).
\end{align*}
\]

For the remaining property \( P_8^* \), suppose that \( \delta > 0 \) and \( \rho(f, g) < \delta \). Then
\( g(x) - \delta < f(x) < g(x) + \delta \) for every \( x \in X \). Since \( \Delta \) is uniformly continuous,
given \( \epsilon > 0 \), there exists a single \( \delta > 0 \) such that
\[
\Delta(g(x)) - \epsilon < \Delta(f(x)) < \Delta(g(x)) + \epsilon
\]
for every \( x \in X \) whenever \( \rho(f, g) < \delta \). In that case

\[
\int [\Delta(g(x)) - \epsilon] \, d\mu(x) < \int \Delta(f(x)) \, d\mu(x) < \int [\Delta(g(x)) + \epsilon] \, d\mu(x),
\]
i.e., \( d(g) - \epsilon < d(f) < d(g) + \epsilon \). The theorem follows.

The above theorem shows that every appropriate function \( \Delta \) on \( I \) gives rise to a degree function \( d(f) = d\Delta(f) \), and in particular this applies to the finite case considered in [2]. Further, if properties \( P_1^* \) to \( P_8^* \) are accepted as formalizations of intuitive stipulations regarding a measure of fuzziness, then they show that this measure \( d(f) \) is unique up to a choice of function \( \Delta \). In [2], the rôle of \( \Delta \) was taken by the particular function

\[
S(\alpha) = -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) \quad [\alpha \in I],
\]

but without further considerations there seem to be no \textit{a priori} reasons for this choice, other than for its formal connection with the probabilistic entropy function. In fact, the simplest function to choose is perhaps the function \( \Delta_0 \) defined by

\[
\Delta_0(\alpha) = \begin{cases} 
\alpha & \text{for } \alpha \in [0, \frac{1}{2}], \\
1 - \alpha & \text{for } \alpha \in [\frac{1}{2}, 1].
\end{cases}
\]

In general, one might wish to choose \( \Delta \) in accordance with the requirements of some particular line of investigation, e.g., in terms of suitable properties of its rates of increase and decrease, perhaps. It is worth emphasizing, however, that apart from indicating the choice available in this direction the present discussion is \textit{not} concerned with the question of obtaining a minimal set of independent axioms characterizing \( d(f) \).

Lastly one may consider the question whether properties \( P_1^* \) to \( P_8^* \) are necessary consequences of the concept of degree of fuzziness. Since this is a problem of the formal interpretation of intuitive ideas, it cannot of course expect a unique solution. However, it is worth noting that at least some modifications that come readily to mind can still be incorporated within the present framework. For example, the authors of [1] consider the possibility that, given a fuzzy set \( f \), one may wish to select two levels \( \epsilon_1 \) and \( \epsilon_2 \) in \( I \) and stipulate that (i) a point \( x \in X \) "belongs" to the fuzzy set if \( f(x) \geq 1 - \epsilon_1 \), (ii) \( x \) "does not belong" if \( f(x) \leq \epsilon_2 \), and (iii) \( x \) is "indeterminate" if \( \epsilon_2 < f(x) < 1 - \epsilon_1 \). There are obviously many variations to this type of convention, and here the approach to a degree of fuzziness via \( P_1^* \) to \( P_8^* \) is not directly suitable in all respects. Nevertheless, instead of seeking appropriate modifications of these properties directly, one may often derive equivalent conclusions by retaining \( P_1^* \) to \( P_8^* \) but associating a modified fuzzy set \( f' \) with \( f \). For example, in the case mentioned above, the authors
of [1] associate with \( f \) the three-valued function \( f' \) such that \( f'(x) = 1 \) if \( x \) "belongs" to \( f \), \( f'(x) = 0 \) if \( x \) "does not belong", and \( f'(x) = \frac{1}{2} \) if \( x \) is indeterminate. In this case, it would often be easy to apply the previous theory of \( d(f) \) to \( f' \) directly, and thus avoid modifying the entire theory for the sake of a particular (nonunique) convention. Similar comments apply to other cases.

\textit{Note added in proof.} After submission of this paper, two articles appeared which discuss some further measures of the type \( d_{\mu}(f) \) when \( X \) is a finite set. The articles are by Capocelli and De Luca (Inform. Contr. 23 (1973), 446-473), and De Luca and Termini (Inform. Contr. 24 (1974), 55-73).

\textbf{REFERENCES}