A New Derivation of Centroid Defuzzification

Mourad Oussalah
Centre for Software Reliability, City University, London, UK, m.oussalah@city.ac.uk

Hung T. Nguyen
Mathematics, New Mexico State University, Las Cruces, NM 88003, USA, hunguyen@nmsu.edu

Vladik Kreinovich
Computer Science, University of Texas, El Paso, TX 79968, USA, vladik@cs.utep.edu

***Abstract*—We describe a new symmetry-based derivation of centroid defuzzification.**

The Need for Defuzzification. Fuzzy logic and fuzzy control start with the knowledge expressed by experts in terms of words from a natural language, and end up with control or decision recommendations; see, e.g., [1], [2], [4].

As a result of the standard fuzzy control methodology, we get a fuzzy set (membership function) $\mu(u)$ which describes, for each possible control value u, how reasonable it is to use this particular value. In automatic control applications, we want to transform this fuzzy recommendation into a single value $\bar{u}$ of the control that will actually be applied. This transformation from a fuzzy set to a (non-fuzzy) number is called a *defuzzification.*

What defuzzification should we apply?

The Standard Choice of a Defuzzification is:

$$\bar{u} = \frac{\int u \cdot \mu(u)\,du}{\int \mu(u)\,du}. \qquad (1)$$

This formula is called *centroid defuzzification*, because it resembles a formula from mechanics that describes the *center of mass* $\vec{r}$ of a system of several points with masses m_i at locations $\vec{r}_i$ as

$$\vec{r} = \frac{\sum m_i \cdot \vec{r}_i}{\sum m_i}; \qquad (2)$$

here, u is the analog of a location, and $\mu(u)$ is the analog of the mass.

This formula has been successfully used in fuzzy control.

Warning: Centroid Defuzzification Is Only a Basis, It Does Not Always Work By Itself. In most real-life situations, centroid defuzzification leads to a meaningful control, but in some cases, it may lead to an unreasonable control.

Let us give a simple example. Suppose that we are designing an automatic controller for a car. If the car is traveling on an empty wide road, and there is an obstacle straight ahead (e.g., a box that fell from a truck), then a reasonable idea is to *swerve* to avoid this obstacle. Since the road is empty, there are two possibilities: we can swerve to the right, we can swerve to the left.

For swerving, the control variable u is the angle to which we steer the wheel. Based on the distance to the obstacle and on the speed of the car, an experienced driver can describe a reasonable amount of steering u_0. (In reality, u_0 will probably be a fuzzy value, but for simplicity, we can assume that u_0 is precisely known.) Thus, as a result of formalizing expert knowledge, we conclude that there are two possible control values: the value u_0 with degree of confidence $\mu(u_0) = 1$, and the value $-u_0$, with degree of confidence $\mu(-u_0) = 1$. If we apply the defuzzification formula (1) to this situation, we get $\bar{u} = 0$. So, the recommended control means that *no swerving* will be applied at all, and the car will run straight into the box.

To avoid such situations, we must *modify* the centroid defuzzification. We could have screened out the value $\bar{u} = 0$ because for this value, $\mu(\bar{u}) = 0$. Thus, instead of using a simply centroid defuzzification, we can do the following:

- First, we apply defuzzification to the original membership function $\mu(u)$.
- Then, we check whether the resulting control value $\bar{u}$ is reasonable, i.e., whether the degree of confidence $\mu(\bar{u})$ is big enough (e.g., larger than some pre-defined value μ_0).
- If $\mu(\bar{u}) \geq \mu_0$, then we apply the control $\bar{u}$.
- If $\mu(\bar{u}) < \mu_0$, this means that there are several areas of reasonable control separated by a gap, and $\bar{u}$ happens to be in this gap. In this case, instead of applying the the centroid defuzzification to the entire *membership* function $\mu(u)$, we select *one of the areas*, and then apply centroid defuzzification only to value u from this area.

This idea was first proposed and successfully implemented by J. Yen [5]–[7].

A Related Probabilistic Derivation of Centroid Defuzzification. The above formula can be naturally reformulated in probabilistic terms. Indeed, according to the standard probabilistic decision making approach, we must select the value $\bar{u}$ for which the average loss $\int(\bar{u} - u)^2 \cdot \rho(u)\,du$ is the smallest possible, where $\rho(u)$ is the probability density characterizing the probability of different values of u. We cannot apply this criterion directly, because we do not know the values $\rho(u)$. Instead, we know, for each value u, the degree $\mu(u)$ to which this value u is possible. We need to "translate" these degrees into a subjective probability distribution $\rho(u)$.

It is reasonable to assume that the larger our degree, the larger the probability. The simplest way to express this assumption would be to assume that $\rho(u) = \mu(u)$, but we cannot do that, because the probability density must be normalized ($\int \rho(u)\,du = 1$), while the membership function may be not normalized at all. Thus, we have to use the next simplest way, by using $\rho(u) = c \cdot \mu(u)$ and selecting the constant c so as to guarantee the normalization.

For this choice of the density function, minimizing the average loss is equivalent to minimizing the expression $\int (\bar{u} - u)^2 \cdot \mu(u)\, du$, which, as one can easily see, leads to centroid defuzzification.

The Above Derivations Were Heuristic; We Need a More Justified Derivation. Most known derivations of centroid justification are heuristic. For example, in the probabilistic justification, we could as well take $\rho(u) = f(\mu(u))$ for some nonlinear monotonic function $f(z)$, and end up with a different expression for defuzzification.

It is known that different versions of fuzzy control methodology lead to different control quality (see, e.g., [3]), so it is important to try our best in selecting this methodology. We would therefore like either to get a more precise (and more confidence-bringing) justification for centroid defuzzification, or, alternatively, to find a better defuzzification procedure.

In this paper, we provide a more justified derivation of centroid defuzzification (thus showing that alternative defuzzification procedures may not be so good). This justification will be done in terms of *invariance*, in the same style in which in [2], we justified different operations and properties of fuzzy logic. Before we describe our result, let us list and motivate the corresponding invariance requirements.

First Invariance: With Respect to Rescaling of Certainty Degrees $\mu(u)$. One of the natural methods to ascribe the degree of confidence $d(A)$ to a statement A is to take several (N) experts, and ask each of them whether he or she believes that A is true. If $N(A)$ of them answer "yes", we take $d(A) = N(A)/N$ as the desired certainty value. If all the experts believe in A, then this value is 1 (=100%), if half of them believe in A, then $t(A) = 0.5$ (50%), etc.

Knowledge engineers want the system to include the knowledge of the entire scientific community, so they ask as many experts as possible. But asking too many experts leads to the following negative phenomenon: when the opinion of the most respected professors, Nobel-prize winners, etc., is known, some less self-confident experts will not be brave enough to express their own opinions, so they will rather say nothing. How does their presence influence the resulting uncertainty value?

Let N denote the initial number of experts, $N(A)$ the number of those of them who believe in A, and M the number of shy experts added. Initially, $d(A) = N(A)/N$. After we add M experts who do not answer anything when asked about A, the number of experts who believe in A is still $N(A)$, but the total number of experts is bigger $(M + N)$. So the new value of the uncertainty ratio is $d'(A) = \dfrac{N(A)}{N+M} = c \cdot d(A)$, where we denoted $c = N/(M+N)$. Thus, the same confidence expressed by a membership function $\mu(u)$ can be alternatively expressed by a membership function $\mu'(u) = c \cdot \mu(u)$.

The result of defuzzification should not change if we simply change the way the same confidence is expressed. In mathematical terms, this requirement means that the defuzzification procedure should be *invariant* with respect to the transformation $\mu(u) \to c \cdot \mu(u)$. To distinguish this invariance requirement from several other invariances which we will introduce later, we will call this invariance *μ-invariance.*

Because of This Invariance, We Cannot Have an Analytical Defuzzification Operation, We Can Only Have a Fractionally Analytical One. It turns out that the above requirement directly affects a mathematical formalization of our problem.

Indeed, a defuzzification operation D should transform a function $\mu(u)$ into a value $\bar{u}$. A function $\mu(u)$ can be viewed as a sequence of its values, so a defuzzification operation, in effect, maps the sequences $(\mu(u_1), \ldots, \mu(u_n))$ into values $\bar{u}$. From this viewpoint, selecting D is equivalent to selecting a function of n variables $\bar{u} = D(\mu(u_1), \ldots, \mu(u_n))$.

It seems natural to consider *analytical* functions, i.e., functions which can be expanded into Taylor series:

$$\bar{u} = a_0 + \sum_{i=1}^{n} a_i \cdot \mu(u_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} \cdot \mu(u_i) \cdot \mu(u_j) + \ldots$$

In reality, there are *infinitely many* possible values u. To get a formula for this realistic case, we must take more and more points and then tend this number of points to infinity. Then, the sums tend to integrals, and we get the following formula:

$$\bar{u} = D(\mu) = a_0 + \int a(u) \cdot \mu(u)\, du +$$

$$\int\int a(u, u') \cdot \mu(u) \cdot \mu(u')\, du\, du' + \ldots \qquad (3)$$

Since the result should not change under the transformation $\mu \to c \cdot \mu$, we should have $D(c \cdot \mu) = D(\mu)$ for an arbitrary constant c. Since the value $D(c \cdot \mu)$ does not depend on c at all, in particular, when $c \to 0$, the value $D(c \cdot \mu)$ tends to exactly the same constant value $D(\mu)$, i.e., we have $D(c \cdot \mu) \to D(\mu)$.

However, if we substitute $c \cdot \mu$ into the expression (3) and tend c to 0, we can easily see that $D(c \cdot \mu) \to a_0$. Thus, for the expression (3), the above invariance leads to a meaningless conclusion that for every membership function μ, the value $D(\mu)$ is equal to the same constant a_0. We want the defuzzification procedure to be *nontrivial* in the sense that $D(\mu)$ should not be equal to the same constant. This result shows that, due to the above invariance requirement, we cannot have an analytical defuzzification operation.

The next natural choice is to have *fractional analytical* functions, i.e., ratios $\bar{D}(u = D(\mu) = D^+(\mu)/D^-(\mu)$, where:

$$D^+(\mu) = a_0 + \int a(u) \cdot \mu(u)\, du +$$

$$\int\int a(u, u') \cdot \mu(u) \cdot \mu(u')\, du\, du' + \ldots \qquad (4a)$$

$$D^-(\mu) = b_0 + \int b(u) \cdot \mu(u)\, du +$$

$$\int\int b(u,u')\cdot\mu(u)\cdot\mu(u')\,du\,du'+\dots \quad (4b)$$

For such expressions, invariance is possible: e.g., centroid defuzzification is of this type and is invariant.

We will restrict ourselves to *non-degenerate* ratios of this type, in which linear terms ($a(u)$ and $b(u)$) in the numerator and in the denominator are not identically 0.

Additional Invariances. The numerical value of the control u changes if we change the unit for measuring control and if we change the starting point. For example, if u is the moment of time at which we have to start the spaceship's engine for a descent, then the numerical value of u depends on what units we use for measuring time (seconds, days, etc.), and what starting point we take (we get different numerical values depending on whether we use astronomical time or time from the beginning of this particular spaceflight).

If we change the unit to a one which is k times smaller, then the numerical value u is replaced by $u\to k\cdot u$. If we change the starting point, we get $u\to u+c$. It is reasonable to require that the defuzzification result do not depend on this selection of units.

How does this change of units influence membership functions? For a change of units, the value u in the new units corresponds to the value u/k in the old units. Thus, in new units, the new membership function expressing the same degrees of confidence is equal to $\mu'(u)=\mu(u/k)$. After applying defuzzification to this new function, we should get the same value as before – but expressed in new units. In other words, we the value $D(\mu(u/k))$ in new units should be equal to $D(\mu(u))$ in the old units, i.e., we should have $D(\mu(u/k))/k=D(\mu(u))$, i.e., equivalently, $D(\mu(u/k))=k\cdot D(\mu(u))$.

Similarly, changing the starting point means replacing the original membership function $\mu(u)$ by a new function $\mu'(u)=\mu(u-c)$. The requirement that the defuzzification result should not depend on this change can be similarly expressed as $D(\mu(u-c))=D(\mu(u))+c$.

Thus, in mathematical terms, we require that $D(\mu(u/k))=k\cdot D(\mu(u))$ (i.e., that D is *scale-invariant*) and that $D(\mu(u-c))=D(\mu(u))+c$ (i.e., that D is *shift-invariant*)

We also want to make sure the operation D is *consistent*, in the sense that in the almost crisp case, when $\mu(u)$ is only different from 0 in an interval $[u^-,u^+]$, we should get $D(\mu)$ within this interval.

It turns out that these requirements uniquely determine the centroid defuzzification:

Theorem. *Centroid defuzzification is the only consistent shift-invariant scale-invariant μ-invariant non-degenerate fractional-analytical defuzzification procedure.*

Proof. Let us assume that D is a consistent shift-invariant scale-invariant μ-invariant non-degenerate fractional-analytical defuzzification procedure, and let us show that D coincides with a centroid.

1. First, let us show that for D, the values of a_0 and b_0 in the expansion (4) are both equal to 0.

We will prove this by reduction to a contradiction. Suppose that at least one of the values a_0 and b_0 is different from 0. We assumed that D is μ-invariant, i.e., that $D(c\cdot\mu)=D(\mu)$ for every constant c. When we substitute $\mu'=c\cdot\mu$ instead of μ into the formula (4) and take $c\to 0$, the numerator tends to a_0 and the denominator tends to b_0. Thus, since at least one of the values a_0 and b_0 is different from 0, we would conclude that the ratio $D(c\cdot\mu)$ tends to a constant limit a_0/b_0 (finite or infinite) which, thus, does not depend on the membership function $\mu(u)$. Due to μ-invariance, this ratio is equal to $D(\mu)$. Thus, we conclude that $D(\mu)$ does not depend on μ at all.

This conclusion contradicts to our assumption that the operation D is consistent and thus, cannot be constant: indeed, for different membership functions μ and μ' located on different non-intersecting intervals, the values $D(\mu)$ and $D(\mu')$ belong to these non-intersecting intervals and are, therefore, different.

This contradiction shows that $a_0=b_0=0$.

2. We have just shown that the expression (4) for $D(\mu)$ does not contain constant terms. Let us now show that D can be expressed in the form (4) with only linear terms present.

Indeed, since $a_0=b_0=0$, due to μ-invariance, for every c, we have $D(\mu)=D(c\cdot\mu)$. Here,

$$D^+(c\cdot\mu)=c\cdot\int a(u)\cdot\mu(u)\,du+$$

$$c^2\cdot\int\int a(u,u')\cdot\mu(u)\cdot\mu(u')\,du\,du'+\dots; \quad (5a)$$

$$D^-(c\cdot\mu)=c\cdot\int b(u)\cdot\mu(u)\,du+$$

$$c^2\cdot\int\int b(u,u')\cdot\mu(u)\cdot\mu(u')\,du\,du'+\dots \quad (5b)$$

If we divide both expressions $D^+(c\cdot\mu)$ and $D^-(c\cdot\mu)$ by C, the value of the ratio $D(c\cdot\mu)$ will not change, so from $D(\mu)=D(c\cdot\mu)$ we conclude that

$$D(\mu)=\frac{\int a(u)\cdot\mu(u)\,du+c\cdot\int\int a(u,u')\dots}{\int b(u)\cdot\mu(u)\,du+c\cdot\int\int b(u,u')\dots}. \quad (6)$$

When $c\to 0$, all the terms in the numerator and in the denominator tend to 0 except for the linear terms. So, in the limit $c\to 0$, the equation (6) leads to the desired formula

$$D(\mu)=\frac{\int a(u)\cdot\mu(u)\,du}{\int b(u)\cdot\mu(u)\,du}. \quad (7)$$

3. Let us now use the scale-invariance property.

Scale-invariance means that $D(\mu(u/k))=k\cdot D(\mu(u))$. Substituting $\mu(u/k)$ into the formula (7), we conclude that

$$D(\mu(u/k))=\frac{\int a(u)\cdot\mu(u/k)\,du}{\int b(u)\cdot\mu(u/k)\,du}. \quad (8)$$

We can somewhat simplify this formula if we introduce an auxiliary variable $u'=u/k$; for this variable, $u=k\cdot u'$,

$du = k \cdot du'$, and so, the formula (8) takes the following form:

$$D(\mu(u/k)) = \frac{k \cdot \int a(k \cdot u') \cdot \mu(u')\, du'}{k \cdot \int b(k \cdot u') \cdot \mu(u')\, du'}. \qquad (9)$$

Dividing both the numerator and the denominator by k, and renaming the integration variable by u, we conclude that:

$$D(\mu(u/k)) = \frac{\int a(k \cdot u) \cdot \mu(u)\, du}{\int b(k \cdot u) \cdot \mu(u)\, du}. \qquad (10)$$

Thus, scale-invariance means that

$$\frac{\int a(k \cdot u) \cdot \mu(u)\, du}{\int b(k \cdot u) \cdot \mu(u)\, du} = \frac{k \cdot \int a(u) \cdot \mu(u)\, du}{\int b(u) \cdot \mu(u)\, du}, \qquad (11)$$

or, equivalently, that

$$\left(\int a(k \cdot u) \cdot \mu(u)\, du\right) \cdot \left(\int b(u) \cdot \mu(u)\, du\right) =$$

$$\left(\int k \cdot a(u) \cdot \mu(u)\, du\right) \cdot \left(\int b(k \cdot u) \cdot \mu(u)\, du\right). \qquad (12)$$

Both sides of the equation (12) represent a quadratic form (in terms of the variable $\mu(u)$) which is a product of two linear forms. This representation is known to be unique modulo a multiplicative constant. Thus, the form related to $a(k \cdot u)$ is either proportional to $b(u)$ or to $a(u)$. For k close to 1, it cannot be proportional to $b(u)$ – otherwise, in the limit $k \to 1$, we would conclude that the ratio $D(\mu)$ is a (trivial) constant. Thus, $a(k \cdot u)$ must be proportional to the form related to $a(u)$. Hence, for every k, there must exist a multiplicative constant $C_1(u)$ for which $a(k \cdot u) = C_1(u) \cdot a(u)$.

A general solution to this functional equation is known (see, e.g., [2]), it is $a(u) = a_1 \cdot u^\alpha$ for some constant c_1 and α. (One way to get this solution is to differentiate both sides of the above functional equation by k and then set $k = 1$; the resulting differential equation leads to this formula.)

Similarly, we get $b(u) = b_1 \cdot u^\beta$ for some constants b_1 and β.

4. Let us now exploit the property of shift-invariance.

For shift-invariance, similar arguments lead to the functional equations $a(u + c) = C_3(c) \cdot (a(u) + c)$ and $b(u + c) = C_4(c) \cdot b(u)$.

For $b(u) = b_1 \cdot u^\beta$, we thus get $(u + c)^\beta = C_4(c) \cdot u^\beta$, i.e., that the ratio $(u + c)^\beta / u^\beta$ should be equal to $C_4(c)$ and thus, do not depend on u at all. This is only possible when $\beta = 0$. Hence, $b(u) = \text{const}$. By dividing both the numerator and the denominator of the expression (1) by this constant, we can get $b(u) = 1$.

The equation about $a(u)$ leads to $(u + c)^\alpha = C_3(c) \cdot (u^\alpha + c)$ for all u and c. This is only true for $\alpha = 1$, so $a(u) = a_1 \cdot u$. Thus,

$$D(\mu) = a_1 \cdot \frac{\int u \cdot \mu(u)\, du}{\int \mu(u)\, du}. \qquad (13)$$

5. To complete the proof, we must show that $a_1 = 1$.

We can prove that $a_1 = 1$ by using the consistency requirement. Indeed, when $\mu(u)$ is different from 0 only on the interval $[u^-, u^+]$, then the centroid is located within the same interval. In particular, for $u^- = 1$ and $u^+ = 1 + \varepsilon$, we conclude that the centroid value $\bar{u}$ is between 1 and $1+\varepsilon$. Since the value (13) – which is obtained by multiplying by a_1 – must be within this same interval, the value a_1 must be positive. Thus, $1 \le a_1 \cdot \bar{u} \le 1 + \varepsilon$, hence,

$$\frac{1}{\bar{u}} \le a_1 \le \frac{1 + \varepsilon}{\bar{u}}. \qquad (14)$$

Since $1 \le \bar{u} \le 1 + \varepsilon$, we thus conclude that

$$\frac{1}{1 + \varepsilon} \le \frac{1}{\bar{u}}$$

and

$$\frac{1 + \varepsilon}{\bar{u}} \le 1 + \varepsilon.$$

Thus, from the inequality (14), we can conclude that

$$\frac{1}{1 + \varepsilon} \le a_1 \le 1 + \varepsilon.$$

When $\varepsilon \to 0$, we conclude that $a_1 = 1$.

The theorem is proven.

Acknowledgments

This work was supported in part by NASA grants NCC5-209 and NCC 2-1232, NSF grants CDA-9522207 and 9710940 Mexico/Conacyt, Air Force Office of Scientific Research grant F49620-00-1-0365, and by Grant No. W-00016 from the U.S.-Czech Science and Technology Joint Fund.

The authors are thankful to anonymous referees for their valuable comments.

References

[1] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic: Theory and Applications*, Prentice Hall, Upper Saddle River, NJ, 1995.

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

[3] H. T. Nguyen and V. Kreinovich, "Methodology of fuzzy control: an introduction", In: H. T. Nguyen and M. Sugeno (eds.), *Fuzzy Systems: Modeling and Control*, Kluwer, Boston, MA, 1998, pp. 19–62.

[4] H. T. Nguyen and E. A. Walker, *First Course in Fuzzy Logic*, CRC Press, Boca Raton, FL, 1999.

[5] J. Yen and N. Pfluger, "Path planning and execution using fuzzy logic", In: *AIAA Guidance, Navigation and Control Conference*, New Orleans, LA, 1991, Vol. 3, pp. 1691–1698.

[6] J. Yen and N. Pfluger, "Designing an adaptive path execution system", *IEEE International Conference on Systems, Man and Cybernetics*, Charlottesville, VA, 1991.

[7] 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.

[8] L. A. Zadeh, "Fuzzy Sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.