On the convergence of HLMS Algorithm

Murillo Javier[1] **Serge Guillaume**[2] **Elizabeth Tapia**[1] **Bulacio Pilar**[1]

[1]CIFASIS-CONICET, Universidad Nacional de Rosario, Argentina.
[2]UMR-Itap, Cemagref, Montpellier, France.

Abstract

In multicriteria decision making, the study of attribute contributions is crucial to attain correct decisions. Fuzzy measures allow a complete description of the joint behavior of attribute subsets. However, the determination of fuzzy measures is often hard. A common way to identify fuzzy measures is HLMS (Heuristic Least Mean Squares) algorithm. In this paper, the convergence of the HLMS algorithm is analyzed. First, we show that the learning rate parameter (α) dominates the convergence of HLMS. Second, we provide an upper bound for α that guarantees HLMS convergence. In addition, a toy example shows the descriptive power of fuzzy measures versus the poverty of individual measures.

Keywords: multicriteria, fuzzy integrals, HLMS, convergence

1. Introduction

In multicriteria decision making [1, 2, 3, 4], values taken by attributes represent the satisfaction degree or attractiveness felt by the decision maker. To achieve a final decision, a proper attribute aggregation is required. The aggregation method may work just on the individual importance of attributes but also on their collective interaction. An often used individual criteria aggregation is the classical weighted arithmetic mean. In some real situations, however, subsets of attributes work together and the individual aggregation allowed by weighted arithmetic mean assumption happens to be rather far from the reality. For this reason, more general aggregation schemes which consider the collective behavior are looked for, the price to pay being that they are often much more complex to use.

In this paper, we are interested in the collective criteria aggregation through the fuzzy integral and in the identification of its fuzzy measures. Notice that for n attributes, $2^n - 2$ coefficients are needed to specify the model. Regarding the alternatives for fuzzy measure identification [8, 5, 6, 7], we will focus on the HLMS algorithm [9] and its convergence. In particular, a study of the HLMS learning rate parameter and its correct setting for HLMS convergence is presented. The HLMS perform an iterative approximation of fuzzy measures through a gradient descent that minimizes the quadratic difference between the decision value and the Choquet integral on attribute values of a training dataset. Thus, the convergence of HLMS depends on the actual setting of the learning rate parameter.

This paper is organized as follows: Section 2 introduces necessary definition on fuzzy measures and integrals. Section 3 explains HLMS step by step. Section 4 contains the proof of convergence of HLMS. Section 5 gives a toy example where HLMS is applied and the descriptive power of fuzzy measures is compared against weighted sums. Finally, Section 6 brings a conclusion about this work.

2. Basic definitions

We present briefly necessary concepts around fuzzy measures and the Choquet integrals restricted to the finite case. Comprehensive treatments of this topic can be found in [10, 11, 12]. Let us consider a set $X = \{x_1, \ldots, x_i, \ldots, x_n\}$ and $\mathcal{P}(X)$ its power set, $|\mathcal{P}(X)|{=}2^n$.

Definition 1 A fuzzy measure [13] μ defined over X is a set function $\mu\colon \mathcal{P}(X) \to [0, 1]$ that verifies the following axioms:

1. $\mu(\emptyset) = 0$
2. $\mu(X) = 1$
3. $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$

A convenient way to represent graphically the fuzzy measures is through a lattice. The application of the function μ to the $\mathcal{P}(X)$ elements defines the nodes in the lattice. The measures associated with nodes are ordered depending on the inclusion of their elements ($\mu_1 \leq \mu_{12}$), and the measure inclusions are represented by edges. Fig. 1 shows a graphical representation of a fuzzy measure for three attributes. The lattice has $n + 1$ levels numbered from 0 (for the level containing $\mu_\varnothing$) to n (for the level containing μ_X). A *path* in the lattice is a sequence of $n+1$ related nodes, with a node in each level. Given a node in level s, its lower neighbors (respectively upper) are the set of related nodes in level $s-1$ (respectively $s+1$).

The training data is organized in a $P_{[m\times(n+1)]}$. matrix, where n is the number of attributes and m is the number of examples; each column represents an attribute (denoted by x_i) and each row represents an example formed by attributes values $(x_1^z, \ldots, x_n^z)$ and its corresponding target value T^z.

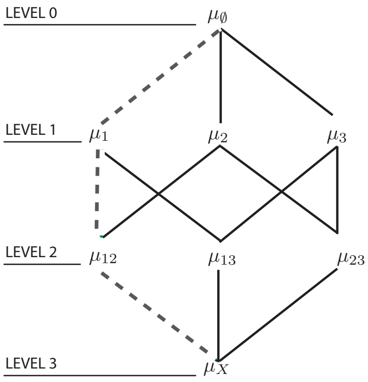

Figure 1: Lattice representation of a fuzzy measure for three attributes

$$P = \begin{pmatrix} x_1^1 & \dots & x_i^1 & \dots & x_n^1 & T^1 \\ \vdots & & \ddots & & \vdots & \\ x_1^z & \dots & x_i^z & \dots & x_n^z & T^z \\ \vdots & & \ddots & & \vdots & \\ x_1^m & \dots & x_i^m & \dots & x_n^m & T^m \end{pmatrix}$$

Definition 2 The discrete Choquet integral [14] of a function $f\colon X \to \Re^+$ with respect to a fuzzy measure μ over X is defined by:
$\mathcal{C}_\mu(f(x_{(1)}), .., f(x_{(n)})) \triangleq \sum_{i=1}^{n}(f(x_{(i)}) - f(x_{(i-1)}))\mu(A_{(i)})$
where n is the number of attributes of P, $A_{(i)} = \{x_{(i)}, x_{(i+1)}, \dots, x_{(n)}\}$, $x_{(\cdot)}$ indicates a reorder of the attributes, i.e., $x_{(1)} < \dots < x_{(n)}$, and $f(x_{(0)}) = 0$.
In our case, X represents a set of attributes and fuzzy measures assess the preference over attribute subsets towards a final decision. The preference relation over alternatives is denoted by $\succ$. For two alternatives P_1 and P_2 , $P_1 \succ P_2$ means that alternative P_1 is preferred to P_2. In multicriteria problems, $\mathcal{C}_\mu(X)$ integrates the individual and collective satisfaction degrees associated with attributes in a unique value which identifies the importance of each example. For the sake of simplicity μ_x denotes $\mu(\{x\})$.

3. HLMS algorithm

The input of HLMS are the rows of P and the output is a vector $S = [\mu_\varnothing, \dots, \mu_X]$ of fuzzy measure values.

Step 0. Initialize the fuzzy measure to the equilibrium state: $\mu_i - \mu_{i-1} = \frac{1}{n}$.

Step 1. Compute the error $e^z = \mathcal{C}_\mu(x^z) - T^z$ considering the data sample (x^z, T^z),. Let denote $u(0), u(1), \dots, u(n)$ the values of the nodes in the *path* involved by x^z. For example, in Fig. 1, we have $u(1) = \mu_1, u(2) = \mu_{12}$. Note that for fuzzy measure definition $u(0) = \mu_\varnothing = 0$ and $u(n) = \mu_X = 1$.

Step 2. Update $u(i)$ as follows :

[1]
$$u^{new}(i) = u^{old}(i) - \alpha \times \frac{e^z}{emax} \times (x^z_{(n-i+1)} - x^z_{(n-i)}) \quad (1)$$

where $\alpha \in [0, 1]$ is called the *learning rate* and $emax$ is the maximum value of error. $emax = 1$ if T^z takes its values in [0,1]. As before, $x^z_{(i)}$ indicates the ith value of X, in ascending order.

Step 3. Check the monotonicity relation. If $e^z > 0$, the verification is done for lower neighbors only, if $e^z < 0$ for upper neighbors only. If a monotonicity relation is violated with $\mu(K)$, then $u(i) = \mu(K)$.

Repeat Step 2 and 3 for $i = 1 \dots (n-1)$ in the following order:
If $e^z > 0$, start with $u(1), u(2), \dots, u(n-1)$
If $e^z < 0$, start with $u(n-1), u(n-1), \dots, u(1)$
Step 1 to Step 3 are repeated for all learning data. This is called one *iteration*.

Step 4. Adjust the value of unmodified nodes in previous steps, considering its upper and lower neighbors. The update is done considering the minimum distance between $\mu(K)$ and its upper (lower) neighbors, denoted by $\overline{d}min$ ($\underline{d}min$) as follows:

$$\mu(K) = \mu(K) + \frac{\overline{d}min - \underline{d}min}{2}$$

4. HLMS convergence analysis

HLMS converges when the difference between the current node value $u^k(i)$ and its real value $u(i)$ decreases after each iteration, i.e., in the $k+1$ iteration the following holds:

$$|u(i) - u^{k+1}(i)| < |u(i) - u^k(i)| \quad (2)$$

HLMS convergence demonstration for a fix training data $(X, \mathcal{C}_\mu(X))$ is presented in this section through the analysis of four possible cases based on Eq.(2).

Case 1: $\boxed{u^k(i) < u(i) \wedge u^{k+1}(i) < u(i)}$ (see Fig.2).

[1] The original update formula was reformulated since an error was committed when derivate the gradient descent criterion $E = \mathcal{C}_\mu(x) - y$.

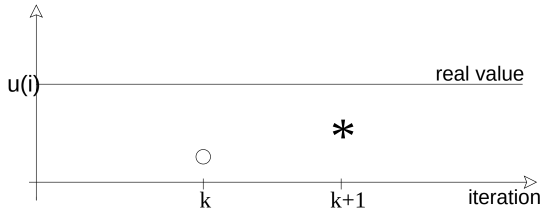

Figure 2: Case 1, both the current $(k+1)$ and previous (k) values are smaller than the real value.

$$
\begin{aligned}
u(i) - u^{k+1}(i) < u(i) - u^k(i) &\Leftrightarrow \\
u^{k+1}(i) > u^k(i) &\Leftrightarrow \\
u^{k+1}(i) - u^k(i) > 0 &\overset{(a)}{\Leftrightarrow} \\
u^k(i) - \alpha(\mathcal{C}_\mu^k(X) - \mathcal{C}_\mu(X))(x_{n-i+1} - x_{n-i}) & \\
-u^k(i) > 0 &\Leftrightarrow \\
\alpha(\mathcal{C}_\mu^k(X) - \mathcal{C}_\mu(X))(x_{n-i+1} - x_{n-i}) < 0 &
\end{aligned}
$$

In (a) the update equation Eq.(1) of Step 2 is applied with e defined in Step 1. Notice that both α and $(x_{n-i+1} - x_{n-i})$ are positive. In addition, Choquet integral is a sum of positive quantities and $u^k(i) < u(i)$. Hence, $\mathcal{C}_\mu^k(X) - \mathcal{C}_\mu(X) < 0$ and the above product is negative verifying the Case 1 convergence.

Case 2: $\boxed{u^k(i) > u(i) \wedge u^{k+1}(i) < u(i)}$ (see Fig.3).

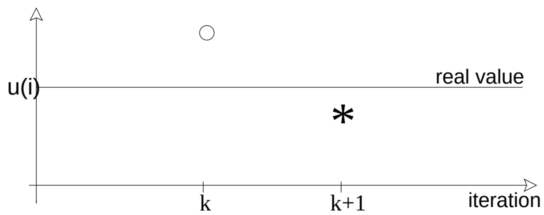

Figure 3: Case 2, the current $(k+1)$ value is smaller than the real value and previous (k) value is bigger than the real value.

$$
\begin{aligned}
u(i) - u^{k+1}(i) < u^k(i) - u(i) &\Leftrightarrow \\
2u(i) - u^{k+1}(i) - u^k(i) < 0 &\overset{(b)}{\Leftrightarrow} \\
2u(i) + \alpha(\mathcal{C}_\mu^k(X) - \mathcal{C}_\mu(X))(x_{n-i+1} - x_{n-i}) & \\
-2u^k(i) < 0 &\Leftrightarrow
\end{aligned}
$$

$$
\alpha < \frac{2(u^k(i) - u(i))}{(\mathcal{C}_\mu(X) - \mathcal{C}_\mu^k(X))(x_{n-i+1} - x_{n-i})} \tag{3}
$$

Similar to Case 1, in (b) the Eq.(1) is applied. We have an upper bound of α that guarantee the HLMS convergence. If α is chosen to satisfy Eq.(3), Case 2 converge.

Case 3: $\boxed{u^k(i) < u(i) \wedge u^{k+1}(i) > u(i)}$ (see Fig.4).

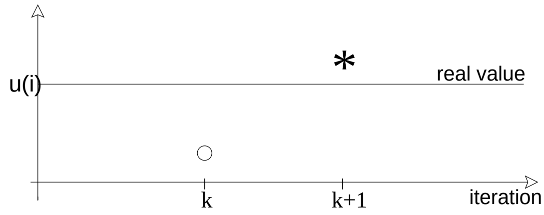

Figure 4: Case 3, the current $(k+1)$ value is bigger than the real value and previous (k) values was smaller than the real value.

$$
\begin{aligned}
u^{k+1}(i) - u(i) < u(i) - u^k(i) &\Leftrightarrow \\
u^{k+1}(i) - 2u(i) + u^k(i) < 0 &\Leftrightarrow \\
2u^k(i) - \alpha \cdot (\mathcal{C}_\mu^k(X) - \mathcal{C}_\mu(X)) \cdot (x_{n-i+1} - x_{n-i}) & \\
-2u(i) < 0 &\Leftrightarrow \\
\alpha < \frac{2(u(i) - u^k(i))}{(\mathcal{C}_\mu(X) - \mathcal{C}_\mu^k(X)) \cdot (x_{n-i+1} - x_{n-i})} &
\end{aligned}
$$

Analogous to the analysis of the previous case, we have an upper bound of α which guarantee the HLMS convergence.

Case 4: $\boxed{u^k(i) > u(i) \wedge u^{k+1}(i) > u(i)}$ (see Fig.5).

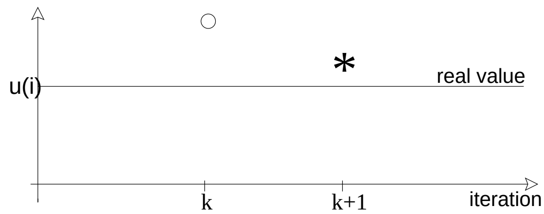

Figure 5: Case 4, the current $(k+1)$ and previous (k) values are bigger than the real value

$$
\begin{aligned}
u^{k+1}(i) - u(i) < u^k(i) - u(i) &\Leftrightarrow \\
u^{k+1}(i) < u^k(i) &\Leftrightarrow \\
u^{k+1}(i) - u^k(i) < 0 &\Leftrightarrow \\
u^k(i) - \alpha \cdot (\mathcal{C}_\mu^k(X) - \mathcal{C}_\mu(X)) \cdot (x_{n-i+1} - x_{n-i}) & \\
-u^k(i) < 0 &\Leftrightarrow \\
\alpha(\mathcal{C}_\mu^k(X) - \mathcal{C}_\mu(X))(x_{n-i+1} - x_{n-i}) > 0 &
\end{aligned}
$$

Both α and $(x_{n-i+1} - x_{n-i})$ are positive. Following a similar analysis to Case 1, $\mathcal{C}_\mu^k(X) - \mathcal{C}_\mu(X) > 0$ and the above product is positive.

5. Toy example

The university wants to buy a computer for the science department. Three factors are considered: microprocessor speed, ram memory and hard disc capacity. Conditions about the computer are:
c1) the most important is microprocessor speed. In this case, a balance between ram memory and hard disc is desired.
c2) when the microprocessor is not quick enough, computers with more ram memory are preferred to improve its performance.
Four alternatives of computers are evaluated and data is organized in Table 1

	Micro(m)	Ram(r)	H. disc(d)
M1	96	78	42
M2	96	66	54
M3	36	78	42
M4	36	66	54

Table 1: Computer acquisition example.

Notice that $M1 \succ M3$ and $M2 \succ M4$.

5.1. Solving with weighted sum we have:

In this case the problem consists in finding the weights, $w()$, of the three attributes according to the preference relations.
Due to condition c1) $M2 \succ M1$

$$\begin{aligned} 96 * w(m) + 78 * w(r) + 42 * w(d) < \\ 96 * w(m) + 66 * w(r) + 54 * w(d) \\ \therefore w(r) < w(d) \end{aligned}$$

and due to c2) $M3 \succ M4$

$$\begin{aligned} 36 * w(m) + 78 * w(r) + 42 * w(d) > \\ 36 * w(m) + 66 * w(r) + 54 * w(d) \\ \therefore w(r) > w(d) \end{aligned}$$

As both conditions lead to contradictory relations, the weighted sum cannot be used to model the preferences.

5.2. Solving with Choquet integral:

In this case, the Choquet integral considers the individual and collective importance of attributes. The problems consists in finding the weights, μ , which represent the importance of each group of attributes.
Due to condition c1)

$$\begin{aligned} 42 + (78 - 42) * \mu(m, r) + (96 - 78) * \mu(m) < \\ 54 + (66 - 54) * \mu(m, r) + (96 - 66) * \mu(m) \\ \therefore 2 * \mu(m, r) < 1 + \mu(m) \end{aligned} \quad (4)$$

Due to condition c2)

$$\begin{aligned} 36 + (42 - 36) * \mu(r, d) + (78 - 42) * \mu(r) > \\ 36 + (54 - 36) * \mu(r, d) + (66 - 54) * \mu(r) \\ \therefore 2 * \mu(r) > \mu(d, r) \end{aligned} \quad (5)$$

	μ_0	μ_m	μ_r	μ_d	$\mu_{m,r}$	$\mu_{m,d}$	$\mu_{r,d}$	μ_X
S^1	0	0.6	0.5	0.5	0.7	0.7	0.6	1

Table 2: μ values used to resolve the toy example with Choquet integral

For instance, values of Table 2 fulfill Eq.(4) and Eq.(5). Then solving for each sample of Table 1 we have:

$$\begin{aligned} M1 = 42 + 36 * \mu(m, r) + 18 * \mu(m) \Rightarrow 78 \\ M2 = 54 + 12 * \mu(m, r) + 30 * \mu(m) \Rightarrow 80.4 \\ M3 = 36 + 6 * \mu(r, d) + 36 * \mu(r) \Rightarrow 57.6 \\ M4 = 36 + 18 * \mu(d, r) + 12 * \mu(r) \Rightarrow 52.8 \end{aligned} \quad (6)$$

As all the conditions are fulfilled, the Choquet integral proves to be able to model these preferences.

5.3. Solving with HLMS:

The HLMS output, μ values, for 300 iterations using adaptive learning rate with initial $\alpha = 0.001$, and target values of Eq.(6) is shown in Table 3.

	μ_0	μ_m	μ_r	μ_d	$\mu_{m,r}$	$\mu_{m,d}$	$\mu_{r,d}$	μ_X
S^1	0	0.6	0.5	0.3	0.7	0.8	0.6	1

Table 3: Output of HLMS for the toy example.

This example only uses *paths* $\{\mu_\varnothing \to \mu_m \to \mu_{m,r} \to \mu_X\}$ and $\{\mu_\varnothing \to \mu_r \to \mu_{r,d} \to \mu_X\}$. Then, μ_d and $\mu_{m,d}$ are computed just on the step 4. Notice that HLMS converges, see Table 3. From the previous analysis we notice that the learning rate α determines the convergence of HLMS, i.e., small α values (Eq.(3)) may help HLMS convergence. However, to guarantee the HLMS convergence, an adaptive learning rate is required.

In Fig.(6) a representation of the convergence behavior of HLMS is shown for the toy example. In the vertical axe the difference between the target and the output of Choquet integral is shown. In Fig.(6) (a) an α value of 0.1 was used with no adaptive learning rate and consequently the HLMS cannot converge. In Fig.(6)(b) an α value of 0.1 is used but, in this case, an adaptive learning rate is used decreasing every 100 iterations. As previously showed, there exists a threshold below which HLMS converge. There is a trade-off between α value, the number of iterations after which α value is decreased and the total number of iterations. A large α value entails fast convergence but unstable. A small α

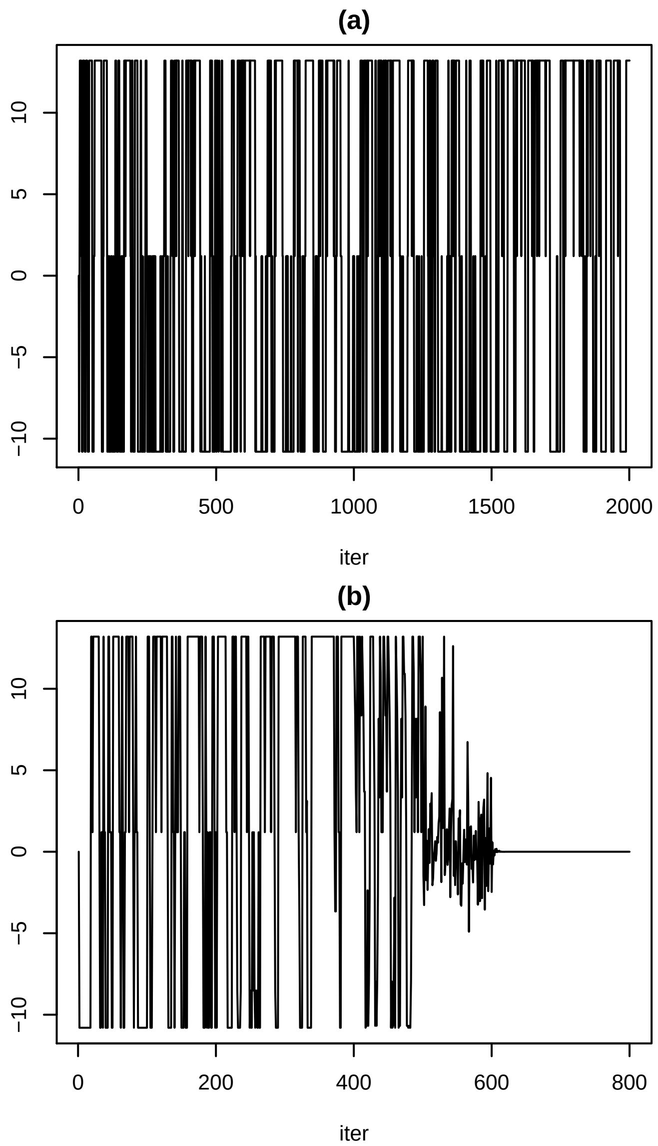

Figure 6: Convergence and divergence of HLMS for the example in section 6

value entails a large numbers of iterations, but due to computer floating point finite accuracy, the algorithm could not converge.

6. Conclusions

In this work we present a convergence analysis of HLMS algorithm. We show that a proper α value allows the HLMS convergence. The bound of α depends on the difference between the real value and the current value. In a convergent HLMS, this difference should be smaller after each iteration. Our results suggest that adaptive learning rates may be best suited to accomplish HLMS convergence. Practical examples showing the descriptive power of fuzzy measures against conventional weighted sum approaches are also analyzed.

References

[1] M. Grabisch. The application of fuzzy integrals in multicriteria decision making. European J. of Operational Research, 89, pages 445-456, 1996.

[2] S.H. Kwon and M. Sugeno. A hierarchical subjective evaluation model using non-monotonic fuzzy measures and the Choquet integral. In M. Grabisch, T. Murofushi, and M. Sugeno, editors, Fuzzy Measures and Integrals - Theory and Applications, pages 375-391. Physica Verlag, 2000.

[3] M. Grabisch. Fuzzy integral in multicriteria decision making. Fuzzy Sets and Systems 69, pages 279-298, 1995.

[4] D. Bouyssou and M. Pirlot. Conjoint measurement tools for MCDM: A brief introduction. In J. Figuera, S. Greco, and M. Ehrgott, editors, Multiple Criteria Decision Analysis: State of the Art Surveys, pages 73-132. Springer, 2005.

[5] Alavi, S. H. and Jassbi, J. and Serra, P. J. and Ribeiro, Rita. Comparison of Genetic and Gradient Descent Algorithms for determining fuzzy measures. IEEE 10th International Conference on Intelligent Enginering Systems (INES 06), London (UK), 2006.

[6] Amol Wagholikar, Peter Deer. Fuzzy Measures Acquisition Methods. Engineering Letters 14(2), pages 56-60, 2007.

[7] Kojadinovic, I.. Estimation of the weights of interacting criteria from the set of profiles by means of information-theoretic functionals. European Journal of Operational Research 155, pages 741-751, 2004.

[8] M. Grabisch, I. Kojadinovic and P. Meyer. A review of methods for capacity identification in Choquet integral based multi-attribute utility theory: Applications of the Kappalab R package, European Journal of Operational Research 186:2, pages 766-785, 2008.

[9] M. Grabisch, A new algorithm for identifying fuzzy measures and its application to pattern recognition, Proceedings of international joint conference of the fourth IEEE international conference on fuzzy systems and the second international fuzzy engineering symposium, Yokohama, Japan, pages 145-150, March 1995.

[10] D. Denneberg. Non-Additive Measure and Integral. Kluwer Academic, 1994.

[11] M. Grabisch, H.T. Nguyen, and E.A. Walker. Fundamentals of Uncertainty Calculi, with Applications to Fuzzy Inference. Kluwer Academic, 1995.

[12] Z. Wang and G.J. Klir. Fuzzy measure theory. Plenum, 1992.

[13] M. Sugeno, Theory of fuzzy integrals and its applications, Doct. Thesis, Tokio Institute of Technology, 1974.

[14] G. Choquet, Theory of capacities, Annales de I'Institut Fourier, Vol 5, pages 131-295, 1953.