International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems
Vol. 0, No. 0 (1993) 000—000

OPTIMIZATION ISSUES FOR FUZZY MEASURES

PEDRO MIRANDA
Universidad de Oviedo
Calvo Sotelo, s/n. 33007 Oviedo, Spain
est010@pinon.ccu.uniovi.es

MICHEL GRABISCH
Thomson-CSF-LCR
Domaine de Corbeville, 91404 Orsay, France
grabisch@lcr.thomson-csf.com

Received (received date)
Revised (revised date)

In this paper, we address the problem of identification of fuzzy measures through different representations, namely the Möbius, the Shapley and the Banzhaf interaction representations. In the first part of the paper, we recall the main results concerning these representations, and give a simple algorithm to compute them. Then we determine the bounds of the Möbius and the interaction representations for fuzzy measures. Lastly, the identification of fuzzy measures by minimizing a quadratic error criterion is addressed. We give expressions of the quadratic program for all the considered representations, and study the uniqueness of the solution.

Keywords: Fuzzy measures, Choquet integral, k-additive measures, learning, uniqueness.

1. Introduction

Recently, the analysis and use of fuzzy measures have been enriched by different equivalent representations of a fuzzy measure,[4] that are obtained through invertible linear transformations applied on the fuzzy measure. The most important ones are the Möbius, the Shapley interaction and the Banzhaf interaction representations. A difficult problem occurring in the practical use of fuzzy measures (e.g. classification[5] or multicriteria decision making[3]), has always been the identification of fuzzy measures from learning data. It is known that the minimization of a squared error criterion leads to a quadratic program, but usually solutions found by this means appear to be unsatisfactory and rather counter-intuitive.

We believe that performing the optimization through either the Möbius or the interaction representation should lead to better results. The reason is that these representations are more meaningful that the usual one, so it is easier to control how the solution is obtained.

The paper presents the different expressions of the quadratic program with respect to the representations. In the last part, we elaborate about the uniqueness of the solution.

Throughout the paper, X will denote an index set of n elements $X = \{1, ..., n\}$.

2. Basic concepts

2.1. *Definitions*

In this section we deal with basic definitions that we will need all over the paper.

Definition 1 *A (discrete)* **fuzzy measure**[9] *on* X *is a set function* $\mu : \mathcal{P}(X) \longrightarrow [0,1]$ *satisfying*

(i) $\mu(\emptyset) = 0, \quad \mu(X) = 1.$

(ii) $A \subset B$ *implies* $\mu(A) \leq \mu(B)$ *(monotonicity).*

A fuzzy measure needs 2^n coefficients to be defined, which are the values of μ for all the different subsets of X. However, we have two of these values fixed, namely $\mu(\emptyset) = 0$, $\mu(X) = 1$.

Definition 2 *A* **pseudo-Boolean function** *is a real valued function* $f : \{0,1\}^n \longrightarrow \mathbb{R}$.

If we identify any subset A of X wiht a point $(x_1, ..., x_n)$ in $\{0,1\}^n$ defined by $x_i = 1$ iff $i \in A$, it is clear that fuzzy measures can be considered as a particular case of pseudo-boolean functions.

It can be shown that any pseudo-Boolean function can be put under a multilinear polynomial of n variables. If $x = (x_1, ..., x_n)$ then

$$f(x) = \sum_{T \subset X} [a_T \prod_{i \in T} x_i], \forall x \in \{0,1\}^n, \tag{1}$$

and with a_T being a real number, $\forall T \subset X$. In fact, a_T coincides with the Möbius transform.

Definition 3 *Let* μ *be a fuzzy measure on X. The* **Möbius transformation** *of* μ *is defined by*

$$a(T) := \sum_{K \subset T} (-1)^{|T \setminus K|} \mu(K), \forall T \subset X. \tag{2}$$

Definition 4 *Let* μ *be a fuzzy measure on X. The* **Shapley interaction index**[4] *of* μ *is defined by*

$$I(T) := \sum_{K \subset X \setminus T} \xi_k^{|T|} \sum_{L \subset T} (-1)^{|T|-|L|} \mu(L \cup K) \tag{3}$$

with $\xi_k^p := \frac{(n-k-p)!k!}{(n-p+1)!}$.

Definition 5 *Let* μ *be a fuzzy measure on X. The* **Banzhaf interaction index** *for any subset* $T \subset X$ *is defined by*

$$J(T) := \frac{1}{2^{n-|T|}} \sum_{K \subset X \setminus T} \sum_{L \subset T} (-1)^{|T|-|L|} \mu(L \cup K). \tag{4}$$

Note that $I(\emptyset)$ and $J(\emptyset)$ are not 0 in general.

Definition 6 *A fuzzy measure μ is said to be* **k-additive** *if its Möbius transformation satisfies $a(A) = 0$ if $|A| > k$ and there exist at least a subset B such that $|B| = k$ and $a(B) \neq 0$.*

Definition 7 *The* **Choquet integral** *of a measurable function $f : X \to \mathbb{R}^+$ with respect to a fuzzy measure μ is defined by*

$$(C)\int f d\mu := \int_0^\infty \mu(\{x|f(x) > \alpha\})d\alpha. \quad (5)$$

When X is finite as in our case, the expression reduces to:

$$(C)\int f d\mu := \sum_{i=1}^{n}(f_{(i)} - f_{(i-1)})\mu(A_i) \quad (6)$$

where f_i stands for $f(i)$, and where parentheses mean a permutation such that $0 = f_{(0)} \leq f_{(1)} \leq \ldots \leq f_{(n)}$ and $A_i = \{(i), \ldots, (n)\}$.

We will denote the Choquet integral in the finite case by

$$\mathcal{C}_\mu(f_1, \ldots, f_n) := (C)\int f d\mu. \quad (7)$$

Chateauneuf and Jaffray[1] have established an equivalent expression for the Choquet integral in terms of the Möbius representation. This expression is:

$$\mathcal{C}_a(f_1, \ldots, f_n) = \sum_{K \subset X} a(K) \min_{i \in K}\{f_i\}. \quad (8)$$

2.2. *Different representations of a fuzzy measure*

It has been proved that a fuzzy measure is uniquely characterized by its Möbius transformation or its Shapley interaction or its Banzhaf interaction. For this reason, we call them *representation* of a fuzzy measure.

Proposition 1 *Let μ be a fuzzy measure, and let a, I and J be its Möbius transformation, its Shapley interaction and its Banzhaf interaction respectively. Then the following holds:*

$$\begin{aligned}
\mu(T) &= \sum_{S \subset T} a(S), \forall T \subset X & (9)\\
I(T) &= \sum_{K \supset T} \frac{1}{|K\backslash T| + 1} a(K) & (10)\\
J(T) &= \sum_{K \supset T} \frac{1}{2^{|K\backslash T|}} a(K) & (11)\\
a(T) &= \sum_{K \subset X\backslash T} B_{|K|} I(K \cup T) & (12)
\end{aligned}$$

$$\mu(A) = \sum_{B\subset X} \beta^{|B|}_{|A\cap B|} I(B) \tag{13}$$

$$a(T) = \sum_{K\supset T} (-\frac{1}{2})^{|K\setminus T|} J(K). \tag{14}$$

where B_k are the Bernoulli numbers which can be computed recurrently through

$$B_k = -\sum_{l=0}^{k-1} \frac{B_l}{k-l+1}\binom{k}{l} \tag{15}$$

starting from $B_0 = 1$, and the β^l_k are given by

$$\beta^l_k := \sum_{j=0}^{k}\binom{k}{j} B_{l-j}. \tag{16}$$

The proofs of these results[1,4,7] are not restricted to fuzzy measures, but can be applied to any set function vanishing on $\emptyset$.

2.3. *Monotonicity constraints on* a, I *and* J

As there are pseudo-Boolean functions which are not fuzzy measures, it is clear that not all Möbius representations nor Shapley interactions correspond to the transformation of a fuzzy measure, i.e. a monotonic set function. The following can be proven:

Theorem 1 *A set of 2^n coefficients $a(T)$, $T \subset X$ corresponds to the Möbius representation of a fuzzy measure if and only if:*

(i) $a(\emptyset) = 0, \quad \sum_{A\subset X} a(A) = 1.$

(ii) $\sum_{i\epsilon B\subset A} a(B) \geq 0$ *for all* $A \subset X$, *for all* $i \in A$.

Theorem 2 *A set of 2^n coefficients $I(A)$, $A \subset X$ corresponds to the Shapley interaction representation of a fuzzy measure if and only if*

(i) $\sum_{A\subset X} B_{|A|} I(A) = 0.$

(ii) $\sum_{i\in X} I(\{i\}) = 1$

(iii) $\sum_{A\subset X\setminus i} \beta^{|A|}_{|A\cap B|} I(A\cup\{i\}) \geq 0 \;, \quad \forall i \in X, \forall B \subset X\setminus\{i\}.$

These results are due to Chateauneuf and Jaffray,[1] and Grabisch.[4] The following one is due to Roubens.[6]

Theorem 3 *A set of 2^n coefficients $J(A)$, $A \subset X$ corresponds to the Banzhaf interaction representation of a fuzzy measure if and only if*

(i) $\sum_{A\subset X} (-\frac{1}{2})^{|A|} J(A) = 0.$

(ii) $\sum_{A\subset X} (\frac{1}{2})^{|A|} J(A) = 1$

(iii) $\sum_{A\subset X\setminus i}(\frac{1}{2})^{|A|}(-1)^{|A\cap B|}J(A\cup\{i\})\geq 0,\ \forall i\epsilon X,\ \forall B\subset X\setminus\{i\}.$

3. Fractal and cardinality tranformations

We are going to study three families of linear transformations defined by Grabisch and Roubens.[6]

- the *fractal* transformation defined recurrently by:

$$\mathbf{F}_{(1)} := \begin{bmatrix} f_1 & f_2 \\ f_3 & f_4 \end{bmatrix},\ f_i \in \mathrm{I\!R},\ i = 1,2,3,4. \tag{17}$$

$$\mathbf{F}_{(k)} := \begin{bmatrix} f_1\mathbf{F}_{(k-1)} & f_2\mathbf{F}_{(k-1)} \\ f_3\mathbf{F}_{(k-1)} & f_4\mathbf{F}_{(k-1)} \end{bmatrix},\ k = 2,\ldots,n. \tag{18}$$

- the *upper-cardinality* transformation defined recurrently by:

$$\mathbf{C}_{(1)} := \begin{bmatrix} c_0 & c_1 \\ 0 & c_0 \end{bmatrix},\ \mathbf{C}^l_{(1)} := \begin{bmatrix} c_{l-1} & c_l \\ 0 & c_{l-1} \end{bmatrix},\quad l = 2,...,n \tag{19}$$

$$\mathbf{C}_{(2)} := \begin{bmatrix} \mathbf{C}^1_{(1)} & \mathbf{C}^2_{(1)} \\ 0 & \mathbf{C}^1_{(1)} \end{bmatrix},\mathbf{C}^l_{(2)} := \begin{bmatrix} \mathbf{C}^l_{(1)} & \mathbf{C}^{l+1}_{l} \\ 0 & \mathbf{C}^l_{(1)} \end{bmatrix},\quad l = 1,...,n-1 \tag{20}$$

$$\mathbf{C}_{(k)} := \begin{bmatrix} \mathbf{C}^1_{(k-1)} & \mathbf{C}^2_{(k-1)} \\ 0 & \mathbf{C}^1_{(k-1)} \end{bmatrix},\quad k = 2,\ldots,n. \tag{21}$$

- the *lower-cardinality* transformation defined recurrently by:

$$\mathbf{C}_{(1)} := \begin{bmatrix} c_0 & 0 \\ c_1 & c_0 \end{bmatrix},\ \mathbf{C}^l_{(1)} := \begin{bmatrix} c_{l-1} & 0 \\ c_l & c_{l-1} \end{bmatrix},\quad l = 2,...,n \tag{22}$$

$$\mathbf{C}_{(2)} := \begin{bmatrix} \mathbf{C}^1_{(1)} & 0 \\ \mathbf{C}^2_{(1)} & \mathbf{C}^1_{(1)} \end{bmatrix},\mathbf{C}^l_{(2)} := \begin{bmatrix} \mathbf{C}^l_{(1)} & 0 \\ \mathbf{C}^{l+1}_{l} & \mathbf{C}^l_{(1)} \end{bmatrix}\quad l = 1,...,n-1 \tag{23}$$

$$\mathbf{C}_{(k)} := \begin{bmatrix} \mathbf{C}^1_{(k-1)} & 0 \\ \mathbf{C}^2_{(k-1)} & \mathbf{C}^1_{(k-1)} \end{bmatrix},\quad k = 2,\ldots,n. \tag{24}$$

Lower and upper cardinality transformations are uniquely defined by a sequence $c_0,...,c_n$ of real numbers. Given $c_0,...,c_n$ the matrix representing the upper-cardinality transformation is the transpose of the matrix representing the lower-cardinality transformation.

Moreover, if $\mathbf{C}^1$ and $\mathbf{C}^2$ represent two upper-cardinality transformations, the sequence c_k related to $\mathbf{C}^1\circ\mathbf{C}^2$ corresponds to

$$c_k = \sum_{l=0}^{k}\binom{k}{l}c^1_{k-l}c^2_l = \sum_{l=0}^{k}\binom{k}{l}c^2_{k-l}c^1_l,\ \forall k \tag{25}$$

where c^1_k, c^2_k are the sequences for $\mathbf{C}^1$ and $\mathbf{C}^2$ respectively.

If $c_0 \neq 0$ (necessary and sufficient condition for the existence of an inverse transformation) the inverse $\mathbf{C}^{-1}$ of the upper-cardinality transformations $\mathbf{C}$ can be obtained by[2]

$$c_k^{-1} = -\sum_{l=0}^{k-1} \binom{k}{l} c_{k-l} c_l^{-1},\ k = 1, ..., n. \tag{26}$$

It is obvious that these results hold also for lower-cardinality transformation because a lower-cardinality transformation is the transpose of an upper-cardinality transformation.

Let us consider the binary order, which is obtained by coding subsets by integers and arranging them in increasing order, i.e. the order given by $\emptyset, \{1\}, \{2\}, \{1,2\}, \{3\}, ...$ Then it can be shown that[6]

$$\mathbf{I} = \mathbf{C}_n^I \mathbf{a},\ \mathbf{J} = \mathbf{C}_n^J \mathbf{a} \tag{27}$$

where $\mathbf{I}, \mathbf{J}$ and $\mathbf{a}$ represent the value vectors of Shapley interaction, Banzhaf interaction and Möbius transformation respectively, taking the binary order; $\mathbf{C}_n^I$ and $\mathbf{C}_n^J$ are matrices representing lower-cardinality transformations defined by

$$c_i^I = \frac{1}{i+1} \text{ and } c_i^J = \frac{1}{2^{i+1}},\ i = 0, ..., n. \tag{28}$$

respectively. Of course, the inverse transformations are also lower-cardinality transformations.

The Möbius transformation

$$a(T) := \sum_{K \subset T} (-1)^{|T \setminus K|} \mu(K)$$

can be rewritten under the fractal form

$$\mathbf{a} = M_{(n)} \circ \mu$$

where μ is the vector of fuzzy measure values and the basic fractal matrix:

$$M_{(1)} := \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}.$$

Note that M is also a lower-cardinality transformation with $c_i = (-1)^i$. Note that $c_0 \neq 0$, thus there exists an inverse transformation, which is

$$\mu_{(n)} = M_{(n)}^{-1} \circ \mathbf{a}$$

where $M_{(n)}^{-1}$ corresponds to the basic fractal matrix:

$$M_{(1)} := \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}$$

that corresponds to a lower cardinality transformation defined by $c_i = 1, \forall i$.

4. Computations over upper and lower cardinality transformations

Since the lower and upper cardinality transformations are $2^n \times 2^n$ matrices, they are not suitable for an efficient implementation. In this section we are going to give an algorithm to obtain the value of a matrix entry given its coordinates. We consider the binary order defined in section 3. We restrict to the upper-cardinality case; the lower-cardinality case is symmetrical. We need some preliminary results. Let $c_0, ..., c_n$ define an upper-cardinality transformation. Then the following can be easily shown by induction:

Lemma 1 *Suppose* $\mathbf{C}^l_{(k)}$ *is defined, and let* $\mathbf{C}^l_k[ij]$ *denote the element in* $\mathbf{C}^l_{(k)}$ *at coordinates* i, j. *Then* $\mathbf{C}^l_k[ij] = 0$ *if* $\mathbf{C}_{(k)}[ij] = 0$, *and* $\mathbf{C}^l_{(k)}[ij] = c_p$ *if* $\mathbf{C}_{(k)}[ij] = c_{p+l-1}$.

Lemma 2 *In an upper-cardinality transformation matrix* $\mathbf{C}$:

1. *the value of* $\mathbf{C}[ij]$ *is 0 if the set associated with* i *is not contained in the set associated with* j.
2. *the value of* $\mathbf{C}[ij] = c_{|A \setminus B|}$ *if the set* B *associated with* i *is contained in the set* A *associated with* j.

Now it is very easy to make an algorithm that computes the upper-cardinality transformation (resp. the lower-cardinality transformation). In the case of upper-cardinality transformation let us denote by A_i the i-th subset of X in the binary order, and let $\mathbf{y}, \mu$ denote the final and the initial vectors respectively. Then the algorithm is:

for (i=1 **to** i=2^n)
    do {
        $y(A_i) = c_0 \cdot \mu(A_i)$
        for (j=i+1 **to** j=2^n)
            do {
                if($A_i \subset A_j$)
                  $y(A_i) = y(A_i) + c_{|A_j|-|A_i|} \cdot \mu(A_j)$
                }
        }

There exists a fast Möbius algorithm developed by Robert Kennes in [8] with a computational cost of $2^n(n/2 - n/2^n)$ if $|X| = n$. With the above explained method we get a computational cost of $3^n - 2^n$ since a subset of only one element is contained in 2^{n-1} subsets of X, a subset of two elements is contained in 2^{n-2} subsets of X and so on; so that there are

$$n2^{n-1} + \binom{n}{2}2^{n-2} + ... + \binom{n}{n}2^0 = \sum_{k=1}^{n}\binom{n}{k}2^{n-k} =$$

$$\sum_{k=0}^{n}\binom{n}{k}2^{n-k} - 2^n = (1+2)^n - 2^n = 3^n - 2^n \qquad (29)$$

multiplications.

Thus, we get a better result with the fast Möbius algorithm, which is not surprising since it was proved to be optimal. However, with the upper-cardinality (resp. lower-cardinality) transformation we are able to compute more transformations that fast Möbius does, namely the transformation from Möbius to Shapley and its inverse transformations that are not possible with the fast Möbius algorithm. This is because the Kennes algorithm is valid only when $c_i = c_0$, $\forall i$.

5. Lower and upper bounds for Möbius and interactions transforms

Theorem 4 *Let A be a subset of X. Then for every fuzzy measure μ*

$$\begin{aligned} a(A) &\leq \sum_{j=0}^{l_{|A|}} \binom{|A|}{j} (-1)^j \\ &= \binom{|A|-1}{l_{|A|}}, \end{aligned} \tag{30}$$

where $l_{|A|}$ is given by:

(i) $l_{|A|} = \frac{|A|}{2}$ *if* $|A| \equiv 0 (\text{mod} \quad 4)$

(ii) $l_{|A|} = \frac{|A|-1}{2}$ *if* $|A| \equiv 1 (\text{mod} \quad 4)$

(iii) $l_{|A|} = \frac{|A|}{2} - 1$ *if* $|A| \equiv 2 (\text{mod} \quad 4)$

(iv) $l_{|A|} = \frac{|A|-3}{2}$ *or* $l_{|A|} = \frac{|A|+1}{2}$ *if* $|A| \equiv 3 (\text{mod} \quad 4)$.

Proof. Let A be a subset of X. We know that

$$\begin{aligned} a(A) &= \mu(A) - \sum_{i \in A} \mu(A \backslash \{i\}) + \sum_{\{i,j\} \in A} \mu(A \backslash \{i,j\}) \\ & \quad - \cdots + (-1)^{|A|-1} \sum_{i \in A} \mu(\{i\}), \forall A \subset X. \end{aligned} \tag{31}$$

We are going to try to find an upper bound for this expression. We consider the fuzzy measure $\bar{\mu}_{|A|}$ defined by:

$$\bar{\mu}_{|A|}(B) = 1 \text{ if } |B| \geq l_{|A|} \tag{32}$$

$$\bar{\mu}_{|A|}(B) = 0 \text{ if } |B| < l_{|A|} \tag{33}$$

This is clearly a fuzzy measure for all $l_{|A|} \geq 1$. For this measure

$$a(A) = \sum_{j=0}^{l_{|A|}} \binom{|A|}{j} (-1)^j. \tag{34}$$

We can always choose $l_{|A|}$ to be even and such that $\binom{|A|}{l_{|A|}-1} \leq \binom{|A|}{l_{|A|}}$ and $\binom{|A|}{l_{|A|}+1} \geq \binom{|A|}{l_{|A|}+2}$. It is easy to see that the value of $l_{|A|}$ as defined in the theorem is suitable.

We are going to prove that this is the maximum value that $a(A)$ can reach. First, note that for every fuzzy measure μ

$$- \sum_{\substack{B \subset A \\ |B|=|A|-l_{|A|}-1}} \mu(B) + \sum_{\substack{B \subset A \\ |B|=|A|-l_{|A|}-2}} \mu(B) - \cdots \leq 0 \tag{35}$$

because if we take $B \subset A$ such that $|B| = |A| - l_{|A|} - 2$ we can always take $B \cup \{j\} \subset A$ and due to monotonicity $\mu(B \cup \{j\}) \geq \mu(B)$. As $\binom{|A|}{l_{|A|}+1} \geq \binom{|A|}{l_{|A|}+2}$,

$$- \sum_{\substack{B \subset A \\ |B|=|A|-l_{|A|}-1}} \mu(B) + \sum_{\substack{B \subset A \\ |B|=|A|-l_{|A|}-2}} \mu(B) \leq 0. \tag{36}$$

Proceeding in this way for the rest of the summands the result is proven.

Then, the maximum of (34) is reached letting $\mu(B) = 0$, $\forall B$ such that $|B| < l_{|A|}$. Thus we can restrict our study to values $\mu(B)$ with $B \subset A$ and $|B| \geq l_{|A|}$. Let μ be a fuzzy measure. Let us prove that

$$- \sum_{\substack{B \subset A \\ |B|=|A|-1}} \mu(B) + \sum_{\substack{B \subset A \\ |B|=|A|-2}} \mu(B) \leq -\binom{|A|}{1} + \binom{|A|}{2} \tag{37}$$

Consider $B \subset A$, $|B| = |A| - 1$. We take $B \backslash \{i\}$ with some $i \in B$. Then, due to monotonicity $-\mu(B) + \mu(B \backslash \{i\}) \leq 0$. We do this for all $B \subset A$, $|B| = |A| - 1$. We can do it because $\binom{|A|}{1} \leq \binom{|A|}{2}$. Then

$$- \sum_{|B|=|A|-1} [\mu(B) - \mu(B \backslash \{i\})] \leq 0. \tag{38}$$

Now it remains $\binom{|A|}{2} - \binom{|A|}{1}$ subsets B such that $B \subset A$, $|B| = |A| - 2$ that were not used. As $\mu(B) \leq 1$ then (36) becomes clear. Proceeding in this way for the rest of the summands we get

$$\begin{aligned} a(A) &\leq \sum_{j=0}^{l_{|A|}} \binom{|A|}{j} (-1)^j \\ &= (-1)^{l_{|A|}} \binom{|A|-1}{l_{|A|}}. \end{aligned} \tag{39}$$

Now, note that $(-1)^{l_{|A|}} = 1$ in all cases. Thus, we obtain the result.

Note that the bound is reached by $\bar{\mu}_{|A|}$ which depends on $|A|$, so this is the smallest upper bound. □

Now, let us turn to the lower bound. The proof is analogous to the proof of the upper bound but the values of $l_{|A|}$ change. The new values are:

(i) $l_{|A|} = \frac{|A|}{2} - 1$ if $|A| \equiv 0 (\bmod \ 4)$

(ii) $l_{|A|} = \frac{|A|-3}{2}$ or $l_{|A|} = \frac{|A|+1}{2}$ if $|A| \equiv 1 (\bmod \ 4)$

(iii) $l_{|A|} = \frac{|A|}{2}$ if $|A| \equiv 2 (\bmod \ 4)$

(iv) $l_{|A|} = \frac{|A|-1}{2}$ if $|A| \equiv 3 (\bmod \ 4)$.

Thus,

$$a(A) \geq -\binom{|A|-1}{l_{|A|}} \tag{40}$$

With an analogous proof we can obtain

Theorem 5 *Let A be a subset of X. Then for every fuzzy measure μ*

$$\begin{aligned} I(A) &\leq \sum_{j=0}^{l_{|A|}} \binom{|A|}{j} (-1)^j \\ &= \binom{|A|-1}{l_{|A|}} \end{aligned} \tag{41}$$

where $l_{|A|}$ depends of $|A|$ as for Möbius.

This is the same value that we have for Möbius. Thus we have the same upper bound that for Möbius. Of course, this is also true for the lower bound.

Finally, remark that for Banzhaf interaction we will obtain the same bounds.

6. Identification based on learning data

The problem of learning is very important in practice. The fuzzy measure is learned by minimizing some criterion. It has been proved[5] that if we consider the quadratic error criterion the problem reduces to solve the quadratic problem:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{u}^T \mathbf{D}_u \mathbf{u} + \mathbf{\Gamma}_u^T \mathbf{u} \\ \text{under the constraint} & \mathbf{A}_u \mathbf{u} + \mathbf{b}_u \geq \mathbf{0} \end{array} \tag{42}$$

where $\mathbf{u}$ is the vector containing the values of $\mu(A)$ for all $A \subset X$. We are going to consider the same problem for the Möbius transformation. In the sequel we will consider the binary order introduced in section 3. Let us suppose that we are given l values of Choquet integral $e_1, ..., e_l$ and the values $f_1^1, ..., f_n^1, ..., f_1^l, ..., f_n^l$. Our goal is to determine a Möbius transformation a that minimizes

$$P = \sum_{k=1}^{l} (\mathcal{C}_a(f_1^k, ..., f_n^k) - e_k)^2. \tag{43}$$

Of course, we have restrictions over a, namely the restrictions of Theorem 1.

It is important to note that it is possible to find more than one solution.

This problem can be reduced to a quadratic problem with $2^n - 1$ variables (for every Möbius transformation $a(\emptyset) = 0$) and $n2^{n-1} + 1$ constraints, which can be written:

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{a}^T \mathbf{D}_a \mathbf{a} + \mathbf{\Gamma}_a^T \mathbf{a} \\ \text{under the constraints} & \mathbf{f}^T \mathbf{a} = 1 \\ & \mathbf{A}_a \mathbf{a} + \mathbf{b}_a \geq \mathbf{0} \end{array} \tag{44}$$

Let us examine this expression: $\mathbf{f}$ is the vector $(1...1)^T$ corresponding to the restriction $\sum_{A \subset X} a(A) = 1$, and $\mathbf{A}_a$ and $\mathbf{b}_a$ correspond to the restrictions (ii) in Theorem 1. $\mathbf{A}_a$ is a matrix of $n2^{n-1}$ rows. Finally, let us examine $\mathbf{D}_a$ and $\mathbf{\Gamma}_a$. Let us fix k. Then:

$$(\mathcal{C}_\mu(f_1^k, ..., f_n^k) - e_k)^2 = (\sum_{A \subset X} a(A) \check{f}^k(A) - e_k)^2, \tag{45}$$

where $\check{f}^k(A) = \min_{i \in A} f_i^k$. Now we obtain

$$(\sum_{A \subset X} a(A) \check{f}^k(A))^2 - 2e_k \sum_{A \subset X} a(A) \check{f}^k(A) + e_k^2 = \mathbf{a}^T \mathbf{F}_k \mathbf{J} \mathbf{F}_k \mathbf{a} - 2e_k \mathbf{f}_k^T \mathbf{a} + e_k^2 \tag{46}$$

with

$$\mathbf{J} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix}.$$

$\mathbf{J}$ is a $(2^n - 1) \times (2^n - 1)$ matrix. $\mathbf{F}_k$ is a diagonal matrix such that $\mathbf{F}_k(i,i) = \check{f}^k(A)$ where A is the $(i+1)$-th set in binary order (the empty set is the first but $a(\emptyset)$ is not a variable). $\mathbf{f}_k$ is the vector containing the diagonal of $\mathbf{F}_k$. Then

$$\mathbf{D}_a = \sum_{k=1}^{l} \mathbf{F}_k \mathbf{J} \mathbf{F}_k \quad \text{and} \quad \mathbf{\Gamma}_a = -2 \sum_{k=1}^{l} e_k \mathbf{f}_i^T. \tag{47}$$

We can consider the same problem for I. In this case we are going to use the expression we got before and the expression of a as a lower-cardinality transformation of I. We know that $\mathbf{a} = \mathbf{C} * \mathbf{I}$, so that using the result before the objective function can be put as

$$\frac{1}{2}\mathbf{a}^T \mathbf{D}_a \mathbf{a} + \mathbf{\Gamma}_a^T \mathbf{a}. \tag{48}$$

Note that we have to consider a new row and a new column for the empty set, since $I(\emptyset) \neq 0$ in general.

If we merge these expressions the objective function reads

$$\frac{1}{2}\mathbf{I}^T \mathbf{C}^T \mathbf{D}_a \mathbf{C} \mathbf{I} + \mathbf{\Gamma}_a^T \mathbf{C} \mathbf{I}. \tag{49}$$

If we rename $\mathbf{D}_I = \mathbf{C}^T \mathbf{D}_a \mathbf{C}, \mathbf{\Gamma}_I = \mathbf{\Gamma}_a \mathbf{C}$ we obtain

$$\frac{1}{2}\mathbf{I}^T \mathbf{D}_I \mathbf{I} + \mathbf{\Gamma_I}^T \mathbf{I}. \tag{50}$$

In this case the constraints are given by theorem 2. Of course, by the same process we can obtain an analogous expression for J.

Now let us turn to the problem of learning k-additive fuzzy measures. We will suppose that we are given l values of the Choquet integral. We want to get a k-additive measure (with $0 < k < n$) that minimizes the quadratic error. If we consider the Möbius transformation we will have $\binom{n}{1} + \cdots + \binom{n}{k}$ variables as $a(\emptyset) = 0$ for all the Möbius transformations of a fuzzy measure.

We can put this problem under a quadratic problem which can be written

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathbf{a}^T\mathbf{D}\mathbf{a} + \mathbf{\Gamma}^T\mathbf{a} \\ \text{under the constraints} & \mathbf{f}^T\mathbf{a} = \mathbf{1} \\ & \mathbf{A}\mathbf{a} + \mathbf{b} \geq \mathbf{0} \end{array} . \tag{51}$$

The objective function can be found in the same way that for the general case. The number of constraints will be the same as in the general case as is shown in the following.

Proposition 2 *If $k > 1$, the number of constraints for the Möbius transformation of a k-additive fuzzy measure is $n2^{n-1} + 1$.*

Proof. We will make the proof by proving that we can not delete any row of the constraints matrix of the Möbius transformation.

It is easy to see that there are no lines of zeroes in $\mathbf{A}$. Now, let us suppose that there are two lines that take the same values for all the subsets which cardinalities at most k. As we have a zero for all the singletons except for the one the constraint refers to, we know that the rows are referred to the same element that we will denote by j. Let us denote by A and B the two subsets that characterizes the constraints. Let $k \in A \backslash B$. Then

$$\{k, j\} \subset A \Rightarrow 1 \text{ in position corresponding to } \{k, j\}$$

$$\{k, j\} \not\subset B \Rightarrow 0 \text{ in position corresponding to } \{k, j\}$$

but this contradicts the fact that the rows are equal. Thus $A \subset B$. Analogously we can prove that $B \subset A$. Then $A = B$ and we are in the same row. □

If $k = 1$ we have $n + 1$ constraints, one for the first condition of theorem 1 and one for each element of X. We can prove analogous results for I and J.

7. Uniqueness of the solution

With the algorithms explained we will obtain a solution of the learning problem. An important problem that comes out is whether we have only one solution.

It has been proved[4] that the set of Möbius transformations is a convex set. Note that it is also a closed set as all the constraints are closed sets.

Suppose that we are given l values of Choquet integral e_1,...,e_l and f_1^1, ..., f_n^1, ..., f_1^l, ..., f_n^l. With these values we build the matrix Δ_a, which is a $l \times (2^n - 1)$ matrix defined by

$$\Delta_a[i, j] = \min_{k \in A_j} \{f_k^i\} = \check{f}^i(A_j), \tag{52}$$

where A_j is the subset which is in the j-th position if we consider the binary order in $\mathcal{P}(X)$.

Let us denote by $\mathcal{M}$ the set of Möbius transformations on X; as a is determined by $a(A), \forall A \subset X$ we can consider that $\mathcal{M}$ is contained in $\mathbf{R}^{2^n-1}$. In the sequel, we will identify a with the vector $(a(\{1\}), ..., a(X))$. We are looking for a Möbius transformation a_e such that

$$d(\Delta_a a_e, e) = \inf_{a \in \mathcal{M}} \{d(\Delta_a a, e)\}. \tag{53}$$

with $e = (e_1 ... e_l)$ and d the Euclidean distance. Note that $\Delta_a \mathcal{M}$ is a closed convex set because $\mathcal{M}$ is so. Consequently, there will be only one z in $\Delta_a \mathcal{M}$ satisfying

$$d(z, e) = \min_{a \in \mathcal{M}} \{d(\Delta_a a, e)\}. \tag{54}$$

Now our problem is to determine whether the system $\Delta_a a = z$ has only one solution. Note that if we have more than one solution we have an infinite number of solutions as any convex combination of the solutions will be also a solution.

Let us start studying the rank of Δ_a. It is clear that $r(\Delta_a) \leq 2^n - 1$. We can easily find examples for which $r(\Delta_a)$ is exactly $2^n - 1$. For example, if we take as values of f the values $f_i = 1$ if $i \in A$ and $f_i = 0$ if $i \notin A$ and we do this for all $A \subset X$, $A \neq \emptyset$ we will obtain a matrix Δ_a such that $r(\Delta_a) = 2^n - 1$.

If $r(\Delta_a) = 2^n - 1$ we can transform the system $\Delta_a a = z$ into an equivalent system $\Lambda_a a = z$ where Λ_a is an invertible matrix since Δ_a may have more than $2^n - 1$ rows. Thus we have only one solution: $a = \Lambda_a^{-1} z$.

What will happen if $r(\Delta_a) < 2^n - 1$? Let a_z be a Möbius transformation that is a solution of $\Delta_a a = z$. Then we are looking for $a \in \mathcal{M}$ such that

$$\Delta_a a = z = \Delta_a a_z \Leftrightarrow \Delta_a (a - a_z) = 0 \Leftrightarrow a - a_z \in \mathrm{Ker}(\Delta_a). \tag{55}$$

Then if a is a solution it will be $a = a_z + u$ with $u \in \mathrm{Ker}(\Delta_a)$.

Now we have to verify that a is indeed a Möbius transformation of a fuzzy measure, that is, it satisfies constraints of theorem 1. Let $\{u_1, ..., u_p\}$ be a base of $\mathrm{Ker}(\Delta_a)$. Then, if a is another solution, it can be written as $a = a_z + \sum_{i=1}^p \lambda_i u_i$. We have the following constraints.

The first constraint in theorem 1 implies:

$$\sum_{j=1}^{2^n-1} a_j = 1 \Leftrightarrow \sum_{j=1}^{2^n-1} a_{zj} + \sum_{i=1}^{p} \lambda_i \sum_{j=1}^{2^n-1} u_{ij} = 1 \Leftrightarrow \sum_{i=1}^{p} \lambda_i \sum_{j=1}^{2^n-1} u_{ij} = 0. \tag{56}$$

The other constraints can be put under the form $\sum_{j \in \mathcal{K}} a_j \geq 0$, with a suitable definition of $\mathcal{K}$.

Let us take a constraint. If $\sum_{j \in \mathcal{K}} a_{zj} > 0$ we will be able to find coefficients λ_i small enough to verify this constraint.

If $\sum_{j \in \mathcal{K}} a_{zj} = 0$ we will need $\sum_{i=1}^p \lambda_i \sum_{j \in \mathcal{K}} u_{zj} \geq 0$.

Note that we have only one solution if and only if $\lambda_i = 0$, $\forall i$; thus, we only need to solve the following problem:

$$\begin{array}{ll} \text{maximize} & \sum_{i=1}^{p} |\lambda_i| \\ \text{under the constraints} & \sum_{i=1}^{p} \lambda_i \sum_{j=1}^{2^n-1} u_{ij} = 0 \\ & \sum_{i=1}^{p} \lambda_i \sum_{j\in A} u_{zj} \geq 0 \end{array} \tag{57}$$

where $\mathcal{K}$ is such that $\sum_{j\in\mathcal{K}} a_{zj} = 0$. If we have only one solution it must be 0.

Note that this process can be applied to fuzzy measures, Shapley and Banzhaf interaction and k-additive measures because in all these situations the hypotheses (convexity and closed set) hold. The only thing that changes is the number of variables. In the case of fuzzy measures the rank of Δ_u is at most $2^n - 2$. Thus, if the rank of Δ_u is $2^n - 2$ we have only one solution. Now, as a Möbius transformation is determined by a fuzzy measure, we can conclude that we have only one solution if the rank of Δ_a is $2^n - 2$, i.e. if $dim(\mathrm{Ker}(\Delta_a)) = 1$.

Now, we have another question: Which is the smallest value of l which allows us to have only one solution in some situation? The answer is given in the following proposition for fuzzy measures.

Proposition 3 *The smallest value for l is 2.*

Proof. First, let us prove that we can find a situation in which there is only one solution and $l = 2$. Let us take $(f_1^1, ..., f_n^1) = (1, 1 - \frac{1}{n}, 1 - \frac{2}{n}, ..., \frac{1}{n})$ and $e_1 = 1$. Then, the expression of Choquet integral is

$$\frac{1}{n}[\mu(X) + \mu(X\backslash\{n\}) + \cdots + \mu(\{1\})]. \tag{58}$$

To obtain the value 1 we need that all the sets in this expression have a measure of 1. Then $\mu(\{1\}) = 1$ and by monotonicity

$$\mu(A) = 1, \forall A, A \ni 1. \tag{59}$$

For the second data vector we take $(f_1^2, ..., f_n^2) = (\frac{1}{n}, \frac{2}{n}, ..., 1)$ and $e_2 = \frac{1}{n}$. To obtain this value of e_2 we need that $\mu(X\backslash\{1\}) = 0$ and by monotonicity

$$\mu(A) = 0, \forall A \subset X\backslash\{1\}. \tag{60}$$

As every subset of X is either containing 1 or not, this determines uniquely the measure.

Now, let us prove that if $l = 1$ we have always an infinite number of solutions. For every data vector we have a chain $\mathcal{X}$ of subsets $A_{(i)}, n \geq i > 1$ such that $A_{(i)} \subset A_{(i-1)}$ which determines the value of the Choquet integral (see definition 7). Suppose that this chain is given by

$$\{1\} \subset \{1,2\} \subset \cdots \subset X\backslash\{n\}. \tag{61}$$

It is always possible to find a chain $\mathcal{Y}$ such that

$$A \in \mathcal{X}, B \in \mathcal{Y} \Rightarrow A \not\subseteq B, B \not\subseteq A. \tag{62}$$

Take for example

$$\{n\} \subset \{n, n-1\} \subset \cdots \subset X \backslash \{1\} \tag{63}$$

Clearly, the values $\mu(A)$, $A \in \mathcal{Y}$ are not influenced by the values $\mu(A)$, $A \in \mathcal{X}$ and thus we have an infinite number of solutions. □

We may think that if the data are such that all the subsets of X are used we will have only one solution. In this case we would need only $\binom{n}{\frac{n}{2}}$ data vectors if n is an even number and $\binom{n}{\frac{n-1}{2}}$ if n is an odd number[5]. However, this is not true as shown in the following example.

Example 1 *Let us take $n = 3$. Let us take the binary order and consider the 3 data vectors given by*

data vector	*corresponding matrix row in Δ_u*
$(1, 0.5, 0)$	$(0.5, 0, 0.5, 0, 0, 0)$
$(0.5, 0, 1)$	$(0, 0, 0, 0.5, 0.5, 0)$
$(0, 1, 0.5)$	$(0, 0.5, 0, 0, 0, 0.5)$

Now, if we take as solution $e_i = 0.5, \forall i$ we will obtain an optimum when

$$\mu(A) = 0.5, \forall A \subset X, A \neq X, \emptyset$$

and another when

$$\mu(\{1\} = \mu(\{2\} = \mu(\{3\} = 0,$$

$$\mu(\{1,2\}) = \mu(\{1,3\}) = \mu(\{2,3\}) = 1.$$

Thus, we have an infinite number of solutions.

Now, we may think that if we have enough data to explore all possible orderings, that is, at least $n!$ we will have only one solution (note that $n! > 2^n - 2$ if $n > 3$). However, this is not true as is shown in the following example.

Example 2 *Let us take $n = 3$. We consider the six data vectors given by*

$$(1, 0.5, 0), (1, 0, 0.5), (0.5, 1, 0), (0, 1, 0.5),$$

$$(0.5, 0, 1), (0, 0.5, 1).$$

We see that all the orderings are involved. Now, if we take as solution $e_i = 0.5, \forall i$ we will obtain an optimum when

$$\mu(A) = 0.5, \forall A \subset X, A \neq X, \emptyset$$

and another when

$$\mu(\{1\} = \mu(\{2\} = \mu(\{3\} = 0,$$

$$\mu(\{1,2\}) = \mu(\{1,3\}) = \mu(\{2,3\}) = 1.$$

Thus, we have an infinite number of solutions.

In conclusion, there is no relation between the possible orderings and the number of solutions. We have seen that we can get only one solution with two data vectors; however, this is a very special case where the values of Choquet integral are 0 or 1 and so lead to a very extreme fuzzy measure; this implies that a lot of constraints are verified with equality and thus they appear as constraints of (57). This implies that only $\lambda_i = 0$ verifies all the constraints of the problem.

1. A. Chateauneuf and J.Y. Jaffray. Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences*, 17:263–283, 1989.
2. D. Denneberg and M. Grabisch. Interaction transform of set functions over a finite set. *Information Sciences*, to appear.
3. M. Grabisch. Alternative representations of discrete fuzzy measures for decision making. *Int. J. of Uncertainty, Fuzziness, and Knowledge Based Systems*, 5:587–607, 1997.
4. M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.
5. M. Grabisch and J.M. Nicolas. Classification by fuzzy integral — performance and tests. *Fuzzy Sets & Systems, Special Issue on Pattern Recognition*, 65:255–271, 1994.
6. M. Grabisch and M. Roubens. Equivalent representations of a set function with application to decision making. In *6th IEEE Int. Conf. on Fuzzy Systems*, Barcelona, Spain, July 1996.
7. M. Grabisch and M. Roubens. An axiomatic approach of interaction in multicriteria decision making. In *5th Eur. Congr. on Intelligent Techniques and Soft Computing (EUFIT'97)*, pages 81–85, Aachen, Germany, september 1997.
8. R. Kennes. Computational aspects of the Möbius tranformation of a graph. *IEEE Tr. on Systems, Man, and Cybernetics*, 22:201–223, 1992.
9. M. Sugeno. *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology, 1974.