Please cite this paper as: G. Beliakov, M. Gagolewski, S. James,
Hierarchical data fusion processes involving the Möbius representation of capacities,
Fuzzy Sets and Systems 433, 1–21, 2022, doi:10.1016/j.fss.2021.02.006

Hierarchical data fusion processes involving the Möbius representation of capacities

Gleb Beliakov[a], Marek Gagolewski[a,b], Simon James[a]

[a] *School of Information Technology, Deakin University, Geelong, VIC 3220, Australia*
[b] *Warsaw University of Technology, Faculty of Mathematics and Information Science, ul. Koszykowa 75, 00-662 Warsaw, Poland*

Abstract

The use of the Choquet integral in data fusion processes allows for the effective modelling of interactions and dependencies between data features or criteria. Its application requires identification of the defining capacity (also known as fuzzy measure) values. The main limiting factor is the complexity of the underlying parameter learning problem, which grows exponentially in the number of variables. However, in practice we may have expert knowledge regarding which of the subsets of criteria interact with each other, and which groups are independent. In this paper we study hierarchical aggregation processes, architecturally similar to feed-forward neural networks, but which allow for the simplification of the fitting problem both in terms of the number of variables and monotonicity constraints. We note that the Möbius representation lets us identify a number of relationships between the overall fuzzy measure and the data pipeline structure. Included in our findings are simplified fuzzy measures that generalise both k-intolerant and k-interactive capacities.

Keywords: Non-additive measures, Capacities, fuzzy measures, 2-step Choquet integral, aggregation operators, high dimensional data

1. Introduction

Here we consider the theory of fuzzy integrals toward practical aspects of their application as aggregation functions for data fusion and analysis. An aggregation function combines numerical or ordered values into a single output either to provide an overall evaluation of a dataset or to compare sets of inputs and make decisions [2]. What distinguishes fuzzy integrals from commonly used functions such as the weighted arithmetic means and median is that inputs' redundancy and complementary interactions can all be incorporated into the aggregation process. This is achieved through identification of an associated fuzzy measure (also referred to as a capacity [6, 12]), which assigns a weight to each subset of inputs.

The discrete Choquet integral [6, 10], an example of one such fuzzy integral, generalizes the weighted arithmetic mean and the popular ordered weighted averaging (OWA) operators. It hence can be used to provide a type of average such that the effective weight applied to each input depends not only on its source but also on its magnitude relative to the other inputs. As a function applied in classification and multivariate approximation, the Choquet integral has been shown [3, 16, 23] to provide performance that rivals that of neural networks, however with the added advantages of exhibiting some reliability in its behavior due to in-built monotonicity, and interpretability due to well-studied analysis tools such as the Shapley index. The drawback of such flexibility is the need to determine each of the capacity parameters (of which there are $2^n - 1$ when considering n inputs).

A useful compromise proposed in [11, 19] involves limiting interactions to groups of inputs up to size k. By leveraging the fact that Möbius representation of the fuzzy measure results in zeros when interaction is additive, the so-called k-additive fuzzy measures facilitate a drastic reduction in the number of variables required, whilst still allowing a high degree of flexibility. However, whilst the number of variables is indeed reduced, for $k > 2$ the number of linear constraints required is unaffected when learning weights using, e.g., linear or quadratic programming [4, 13].

The relationship between certain hierarchical aggregation structures and fuzzy measures has been investigated in [20, 25]. A number of works by Sugeno, Fujimoto, and Murofushi [22, 25] in the 1990s established the conditions under which a Choquet integral can be decomposed into a Choquet integral of Choquet integrals of subgroups. The two-step integral of Mesiar and Vivona [20] further showed a number of properties that follow with respect to the subgroup fuzzy measures if the outer fuzzy measure is additive. Example hierarchical structures are shown in Fig. 1, where inputs x_i leading in from the left are aggregated by each f_j at the first step according to a full covering (a) or a particular partition (b), and then in the second step used as inputs to be aggregated by F, with f_j and F being Choquet integrals with respect to different fuzzy measures. Note that the (b) case leads to a significant reduction in the number of interconnections between the input variables and the latent ones.

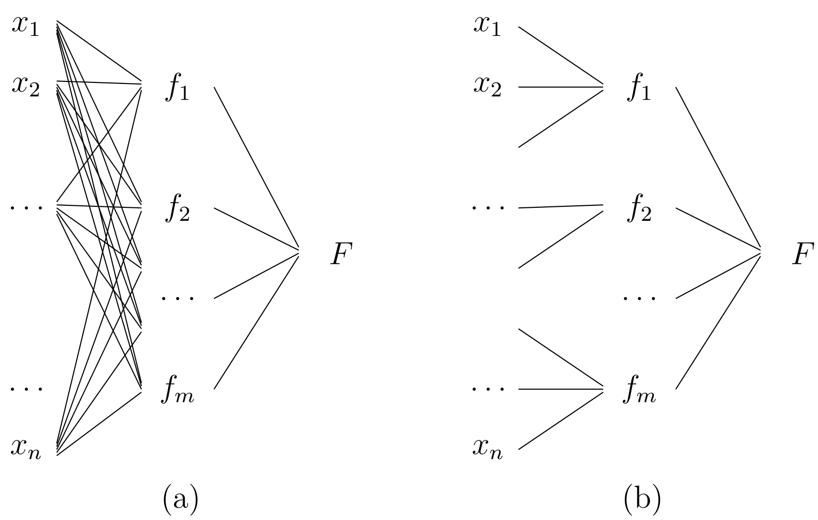


Figure 1: Hierarchical aggregation process involving either a full covering (a) or a partition of the input set (b).

Here we consider hierarchical architectures from the perspective of reducing the number of variables and constraints required in data fitting or other fuzzy measure construction methods. Our goal is to extend the use of the Choquet integral to a much larger number of inputs than before, by exploiting the potential sparsity of fuzzy measure parameters in the Möbius representation. Specific architectures serve as simplifying assumptions that lead to significant reductions, with savings beyond what is even achieved by the k-additive fuzzy measures, hence facilitating huge time savings and tractability for problems involving thousands of variables.

The paper is structured as follows. In the Preliminaries, we give an overview of the background concepts including the Choquet integral, k-additivity, and how parameters can be learnt using linear programming. In Section 3 we examine how the fuzzy measure simplifications referred to as k-intolerance and k-interactivity correspond with particular features in Möbius representation and can also be viewed in the framework of hierarchical aggregation. We also propose a simplified fuzzy measure that encompasses these concepts and refer to it as k-lower/upper interactivity. In Sections 4 and 5 we consider hierarchical aggregation architectures involving two steps (also referred to as two-step Choquet integrals). More precisely, in Section 4 we extend upon some existing results for the case of the second step of aggregation being performed by a Choquet integral with respect to an additive fuzzy measure. We then investigate a number of special cases whereby the second step is with respect to a general fuzzy measure in Section 5. Throughout, our focus is on particular features that can be capitalised upon for practical implementation on larger datasets. We provide indicative tables showing potential savings in terms of the unknown variables and constraints. We conclude in Section 6 with some notes for discussion and future research.

2. Preliminaries

Our results pertain to aggregation performed by the Choquet integral. Here we will give the necessary background on the integral itself, the k-additivity simplification, and how capacity values can be learned from data using linear programming.

2.1. The Choquet integral

When used for the purpose of decision-making and data aggregation, the Choquet integral is best framed in the context of averaging aggregation functions [4].

Definition 1. For an input vector $\mathbf{x} \in [0,1]^n$, a function $f : [0,1]^n \to [0,1]$ is said to be an averaging aggregation function if it is componentwise non-decreasing (or monotone) and idempotent, i.e., $f(t,t,\ldots,t) = t$ for all $t \in [0,1]$.

The monotonicity and idempotency of averaging functions ensure that they remain bounded by the smallest and largest arguments of $\mathbf{x}$, i.e., $\min(\mathbf{x}) \leq f(\mathbf{x}) \leq \max(\mathbf{x})$, which is sometimes referred to as *averaging behavior* or *internality*. Although the class of averaging functions is quite broad, it provides a level of reliability in decision contexts where non-decreasingness may be desirable, and ease of interpretation given that the output will range over the same scale as the inputs.

The weighted arithmetic mean (WAM) is a prototypical example of an averaging aggregation function and one of the most widely adopted in applications. The weights of the arithmetic mean can be interpreted as the level of importance attached to each variable. The Choquet integral generalizes WAM, allowing weights to be assigned not only to each variable but to each coalition thereof. The fuzzy measure is used to encode this information.

Definition 2. For a given finite set $N = \{1, 2, \ldots, n\}$, a fuzzy measure is a set function $\nu : P(N) \to [0, 1]$ defined for all $S \subseteq N$ (the powerset $P(N)$) such that $\nu(\emptyset) = 0, \nu(N) = 1$ and $S \subseteq T$ implies $\nu(S) \leq \nu(T)$.

Definition 3. For a finite set of inputs and a given fuzzy measure ν, the discrete Choquet integral $C_\nu : [0, 1]^n \to [0, 1]$ can then be expressed as

$$C_\nu(\mathbf{x}) = \sum_{j=1}^{n} x_{(j)} \Big(\nu\Big(\{(j), \ldots, (n)\}\Big) - \nu\Big(\{(j + 1), \ldots, (n)\}\Big) \Big),$$

where $(1), \ldots, (n)$ is an ordering permutation of $\mathbf{x}$, i.e., one that yields $x_{(1)} \leq \cdots \leq x_{(n)}$. Moreover, e.g., $\{(j), \ldots, (n)\}$ is the set of indices of the $n - j + 1$ largest values in $\mathbf{x}$.

Calculation of the Choquet integral for a given $\mathbf{x}$ only requires the subsets corresponding to the induced ordering when $\mathbf{x}$ is sorted non-decreasingly. For example, if $\mathbf{x} = (0.5, 0.3, 0.8, 0.1)$, then we would use $\nu(\{1, 2, 3, 4\})$, $\nu(\{1, 2, 3\})$, $\nu(\{1, 3\})$, and $\nu(\{3\})$, with $\nu(\emptyset) = 0$. We can further note from the definition that the function is hence piece-wise linear, with linear behavior on any simplex determined by the ordering of the components of $\mathbf{x}$.

2.2. The k-additivity simplification

An alternative representation of fuzzy measures useful for data fitting and analysis is achieved via the Möbius transform [14].

Definition 4. For a given fuzzy measure ν, we denote the Möbius representation by $\mu : P(N) \to \mathbb{R}$, where

$$\mu_A = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \nu(B).$$

Note that we use μ_A rather than $\mu_\nu(A)$ in order to simplify our notation throughout the rest of the paper. The so-called Zeta-transform, allowing us to convert back from Möbius representation to standard form is given by:

$$\nu(A) = \sum_{B \subseteq A} \mu_B.$$

The Choquet integral in Möbius representation is then

$$C_\nu(\mathbf{x}) = C_\mu(\mathbf{x}) = \sum_{A \subseteq N} \mu_A \min_{i \in A} x_i. \quad (1)$$

For additive fuzzy measures, i.e., where $\nu(A) = \sum_{i \in A} \nu(\{i\})$, the Möbius values for subsets of cardinality 2 or more are all zero. This observation naturally leads to consideration of fuzzy measures with some limited additivity, hence the proposal of k-additive fuzzy measures in [11, 19].

Definition 5. A fuzzy measure is said to be k-additive when for all $A \subseteq N$ such that $|A| > k$, it holds that $\mu_A = 0$.

In fuzzy measure learning and other applications, the k-additivity simplification can greatly reduce the number of variables required to completely define the fuzzy measure. Rather than $2^n - 1$ unknowns, the number of potentially non-zero values in the Möbius representation will be equal to

$$\sum_{i=1}^{k} \frac{n!}{i!(n-i)!}.$$

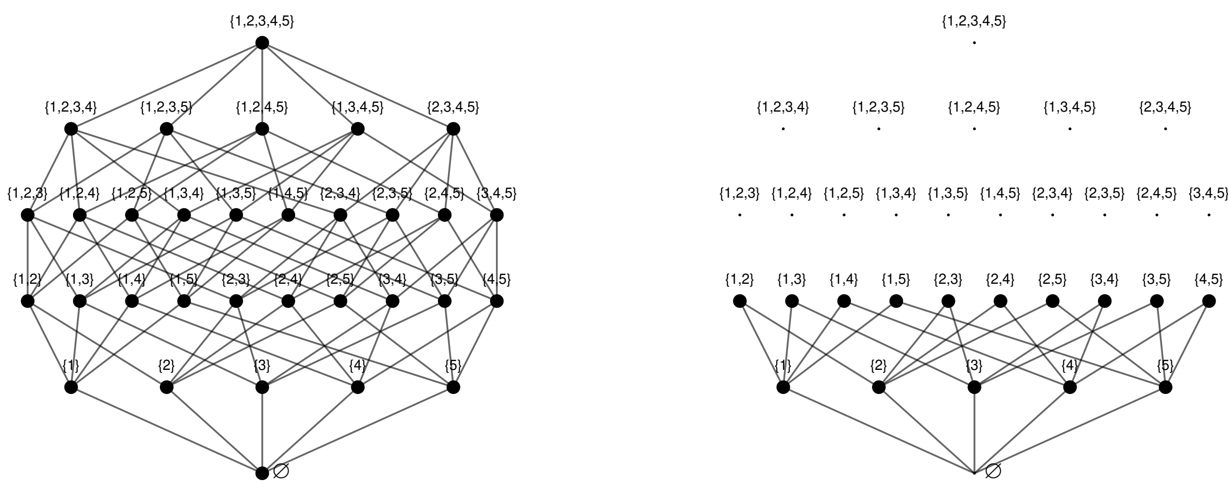
Table 1 gives an idea of the reduction in the number of parameters for some combinations of k and n.

Table 1: Unidentified parameters for k-additive fuzzy measures

	n						
k	2	3	4	5	6	10	100
1	2	3	4	5	6	10	100
2	3	6	10	15	21	55	5050
3		7	14	25	41	175	166750
4			15	30	56	385	$4.09 \cdot 10^6$
5				31	62	637	$7.94 \cdot 10^7$
6					63	847	$1.27 \cdot 10^9$

On a side note, Sugeno fuzzy measures (or λ-fuzzy measures) also allow for a definition by fewer parameters [24], with subsets other than singletons satisfying $\nu(A \cup B) = \nu(A) + \nu(B) + \lambda\nu(A)\nu(B)$. These are somewhat limited in flexibility, however, since the interactions between variables in different subsets will always be captured by λ. Other fuzzy measure simplifications that allow a graduated trade-off in terms of complexity and the number of defining parameters include k-intolerant [17] and k-interactive fuzzy measures [5].

Throughout the paper, we will use visualisations based on the subset relation Hasse diagram to give an impression of the reduction in variables achieved by specific fuzzy measure architectures and how calculations are affected. We will refer to these as *interaction* diagrams. Figure 2(a) shows a standard fuzzy measure defined by $2^n - 1$ variables, alongside Fig. 2(b) a 2-additive fuzzy measure, both for $n = 5$. The smaller nodes correspond with zero-valued subsets in Möbius representation. The larger nodes correspond with non-zero values, and edges are shown from each of these nodes to the subsets with one less cardinality. The edges displayed reinforce the notion of interaction, however for some sparse fuzzy measures the interpretation can be a little more nuanced.



(a) (b)

Figure 2: Interaction diagrams for (a) a full (unconstrained) fuzzy measure and (b) a 2-additive fuzzy measure where edges and nodes for 0-valued subsets (in Möbius representation) are removed.

2.3. Linear programming approach to learning general fuzzy measures

We now look at constructing fuzzy measures from the information provided by experts, which may include their explicit preferences or prototypical cases. Our focus here is not on any particular model but on the complexity of the set of constraints required for consistency with fuzzy measures definition. Consider a dataset consisting of d observed input vectors $\mathbf{x}^{(i)}$ associated with known outputs $y^{(i)}$. The aim is to determine the parameters of our function (i.e., the fuzzy measure weights) such that $f(\mathbf{x}^{(i)})$ is as close as possible to $y^{(i)}$ for all the i across the observed dataset.

We measure closeness using the sum of absolute deviations, i.e.,

$$\sum_{i=1}^{d} |f(\mathbf{x}^{(i)}) - y^{(i)}|$$

and linearize the problem by denoting positive and negative differences respectively by $r_+^{(i)}, r_-^{(i)}$ such that $|f(\mathbf{x}^{(i)}) - y^{(i)}| = r_+^{(i)} + r_-^{(i)}$ and $f(\mathbf{x}^{(i)}) - r_+^{(i)} + r_-^{(i)} = y^{(i)}$, $r_+^{(i)}, r_-^{(i)} \geq 0$.

For each observation $i = 1, \ldots, d$ we have data constraints of the form,

$$\sum_{B \subseteq \{(1),\ldots,(n)\}:(1)\in B} \mu_B x_{(1)}^{(i)} + \sum_{B \subseteq \{(2),\ldots,(n)\}:(2)\in B} \mu_B x_{(2)}^{(i)} + \ldots$$
$$\cdots + \mu_{\{(n)\}} x_{(n)} - r_+^{(i)} + r_-^{(i)} = y^{(i)} \qquad (2)$$

While we have stated the problem of fitting to data in terms of matching the outputs $y^{(i)}$, it is important to mention some alternative approaches. These include those based on classification [18] or ordinal regression, such as the nonadditive robust ordinal regression (NAROR) method [1], which relies on the same variables and constraints setup. There are also quadratic programming formulations and various heuristics (e.g., [13]). Therefore we present the results in a way which is independent of the particular learning strategy or the cost function, by focusing on the constraints which are common to all mentioned approaches.

In Möbius representation, the monotonicity requirement can be expressed as constraints of the following form,

$$\sum_{B \subseteq A: i \in B} \mu_B \geq 0, \qquad (3)$$

which we need for each $i \in A$ and all $A \subseteq N$. The boundary condition $\nu(N) = 1$ can be ensured with the constraint

$$\sum_{A \subseteq N} \mu_A = 1. \qquad (4)$$

While we may have a reduction in variables due to k-additivity, other than the case of $k = 2$, the number of monotonicity constraints in the form of (3) remains. In general, the number of monotonicity constraints required will be,

$$\sum_{i=1}^{n} i\frac{n!}{(i)!(n-i)!} = \sum_{i=1}^{n} \frac{n!}{(i-1)!(n-i)!} = n2^{n-1}.$$

One approach to reducing the number of monotonicity constraints is to use *belief measures*. Belief measures can be defined by what is referred to as a *basic probability assignment*, a set function B defined over the powerset of N satisfying $\mathsf{B}(\emptyset) = 0$, $\mathsf{B}(A) \geq 0$ for all $A \subseteq N$ and $\sum\limits_{A\subseteq N} \mathsf{B}(A) = 1$. The basic probability assignment corresponds with the Möbius representation of a belief measure, and hence the non-negativity condition will automatically ensure monotonicity. This means that all monotonicity constraints except those corresponding with $\mu(A) \geq 0$ for each subset will be redundant. The fitting problem can be further simplified by limiting the number of potential non-zero valued subsets (referred to as *focal elements*). Such measures, however will only exhibit positive interaction effects and hence their use and flexibility in application may be limited.

3. Order-dependent hierarchical aggregation

While the k-additive fuzzy measure simplification allows reduction of variables when fitting Möbius values, there have been other simplifications proposed, which fix the effective weight applied for the lowest or highest k inputs. Such simplifications can be viewed as order-dependent hierarchical aggregation processes and also have notable features in Möbius representation.

In this section we will frequently draw upon the relationship between a fuzzy measure and its dual. For a fuzzy measure in standard representation ν, its dual ν^d is given by

$$\nu^d(A) = 1 - \nu(A'),$$

where A' is the set complement of A. The Choquet integral with respect to a dual fuzzy measure then satisfies $C_{\nu^d}(\mathbf{x}) = 1 - C_\nu(1 - \mathbf{x})$, where $1 - \mathbf{x} = (1 - x_1, 1 - x_2, \ldots, 1 - x_n)$.

It has been shown in [12] that in Möbius form it holds that

$$\mu^d_A = (-1)^{|A|+1} \sum_{B\supseteq A} \mu_B. \tag{5}$$

Eq. (5) tells us that, for each subset A, the Möbius values of the dual fuzzy measure are calculated from the *supersets* $B \supseteq A$.

Remark 6. Eq. (5) makes it easy to see that k-additive fuzzy measures will have duals that are also k-additive, since when $|A| = k$, the only superset B considered for each A will be A itself. Hence, for these sets we have $\mu^d_A = \mu_A$ for odd cardinality and $-\mu_A$ for even cardinality.

We now consider the notions of k-intolerance and k-interactivity in the hierarchical framework, then propose a simplification that generalises both.

3.1. k-tolerance and k-intolerance

With k-intolerance and k-tolerance, we have the following definition.

Definition 7. A fuzzy measure ν is k-intolerant if $\nu(A) = 0$ for all $A \subseteq N$ such that $|A| \leq n - k$ and there exists a B with $|B| = n - k + 1$ such that $\nu(B) \neq 0$. A fuzzy measure is k-tolerant if $\nu(A) = 1$ for all $A \subseteq N$ such that $|A| \geq k$ and there exists a B with $|B| = k - 1$ such that $\nu(B) \neq 1$.

A Choquet integral with respect to a k-intolerant fuzzy measure hence simplifies to an aggregation of the smallest k values only. Given the correspondence between Möbius and standard representation of fuzzy measures, we have the following proposition.

Proposition 1. *A fuzzy measure μ in Möbius representation is k-intolerant if $\mu_A = 0$ for all $A \subseteq N$ such that $|A| \leq n - k$ and there exists a B with $|B| = n - k + 1$ such that $\mu_B \neq 0$. A fuzzy measure is k-tolerant if its dual is k-intolerant.*

Proof. This follows directly from the definition of Möbius representation and k-intolerance (Definitions 7 and 4 respectively). For any $\nu(A) = 0$, all the subsets $B \subseteq A$ will also satisfy $\nu(B) = 0$ and hence $\mu_A = 0$. □

When representing k-intolerant fuzzy measures using the interaction diagrams, we can observe that only the values allocated to the larger subsets are non-zero valued (see Fig. 3). The effect is that the largest $n-k$ inputs are ignored in the aggregation. In Fig. 3(b) it is worth reiterating that the 3-tuple subsets have Möbius values of 0, however the edges are still displayed from the 4-tuples to their subsets in order to give an impression of the variable interactions.

While the k-intolerant fuzzy measures exhibit these prevalent zeros in Möbius representation, this is not the case for the dual k-tolerant fuzzy measures. From Eq. (5), it is apparent that supersets of the dual (the k-intolerant

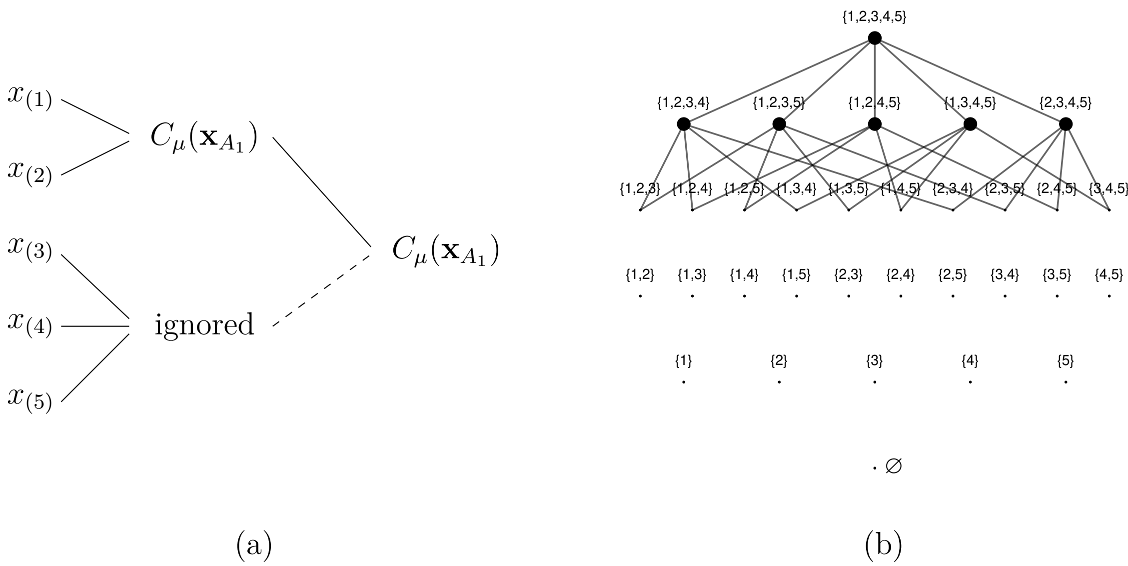


Figure 3: (a) Order-dependent architecture equivalent to a 2-intolerant fuzzy measure - the $x_{(i)}$ are ordered non-decreasingly, and (b) the interaction diagram representing its Möbius representation. All subsets such that $|A| \leq 3$ are set to 0.

fuzzy measure) will not usually be zero-valued. The following example helps illustrate that we would not observe interaction diagrams with similar structures to Fig. 3(b).

Example 1. The fuzzy measure with $\mu_{\{1\}} = \mu_{\{2\}} = \mu_{\{3\}} = 0, \mu_{\{1,2\}} = 0.6, \mu_{\{1,3\}} = 0.3, \mu_{\{2,3\}} = 0.5, \mu_{\{1,2,3\}} = -0.4$, is 2-intolerant. Calculating its 2-tolerant dual gives $\mu_{\{1\}} = 0.5, \mu_{\{2\}} = 0.7, \mu_{\{3\}} = 0.4, \mu_{\{1,2\}} = -0.2, \mu_{\{1,3\}} = 0.1, \mu_{\{2,3\}} = -0.1$, and $\mu_{\{1,2,3\}} = -0.4$.

So whilst k-tolerant fuzzy measures would not directly be obtained from fitting to a reduced number of variables in Möbius representation, we can use the duality relationship and fit to the dual. For a fuzzy measure μ that is k-tolerant, we can learn the corresponding k-intolerant fuzzy measure μ^d by transforming the dataset. In terms of our original fitting objective, if the data take values over the interval $[0, 1]$ we can express the dual fitting objective as

$$\sum_{i=1}^{d} |C_\mu(\mathbf{x}^{(i)}) - y^{(i)}| = \sum_{i=1}^{d} |1 - C_{\mu^d}(1 - \mathbf{x}^{(i)}) - y^{(i)}|. \tag{6}$$

Once the μ^d values have been elicited, Eq. (5) can then be used to obtain μ directly (or the ν representation could also be calculated). We hence are still able to capitalise on the same reductions in the number of variables or monotonicity constraints.

Remark 8. We can make mention of co-Möbius form [12], which also serves as an alternative representation of fuzzy measures and is invertible. For a fuzzy measure ν, the co-Möbius values $\bar{\mu}_A$ for each subset are given by

$$\bar{\mu}_A = \sum_{B \supseteq N \setminus A} (-1)^{|B'|} \nu(B) = \sum_{B \subseteq A} (-1)^{|B|} \nu(B'),$$

where B' is the set complement of B. From Def. 4 and Eq. (5) one can ascertain that for any fuzzy measure with values $\mu_A = 0$, it will hold that $\bar{\mu}^d_A = 0$, i.e., the dual fuzzy measure in co-Möbius representation will have zero values corresponding with the same subsets. We could therefore also fit to the co-Möbius representation, however for simplicity we will continue to focus on Möbius representation.

For k-intolerant fuzzy measures, the number of unknown variables in Möbius representation will be equal to

$$\sum_{i=n-k+1}^{n} \frac{n!}{i!(n-i)!} = \sum_{i=0}^{k-1} \frac{n!}{i!(n-i)!}.$$

In addition to reduction in variables, we also reduce the number of monotonicity constraints required, since clearly all subsets $B \subseteq A$ with $|A| = n-k$ will also be zero too. We hence need only consider

$$\sum_{i=n-k+1}^{n} i \frac{n!}{i!(n-i)!} = \sum_{i=0}^{k-1} (n-i) \frac{n!}{i!(n-i)!}$$

monotonicity constraints.

3.2. k-interactivity

The k-interactive fuzzy measures were proposed in [5]. When framed as a kind of order-dependent hierarchical aggregation process, we see that they capture a similar idea to the k-intolerant fuzzy measures, except that rather than ignoring the smallest (or largest) values, their arithmetic mean is used. This is a consequence of maximising the entropy over the smallest (or largest) inputs.

Definition 9. A fuzzy measure ν is k-interactive if for some chosen $w \in [0,1]$ and $1 \leq k \leq n$,

$$\nu(A) = w + \frac{|A| - k - 1}{n - k - 1}(1 - w), \text{ for all } A, |A| > k.$$

A Choquet integral with respect to a k-interactive fuzzy measure takes a weighted mean (with weights $(1-w)$ and (w) respectively) of the arithmetic mean of the smallest $n - k - 1$ values along with a Choquet integral of the remaining $k + 1$ values.

The relationship between k-intolerance and k-interactivity is most clearly observed by considering the dual of k-interactive fuzzy measures in their Möbius representation.

Proposition 2. *For a given k-interactive fuzzy measure ν expressed in standard representation and satisfying Def. 9 for a given k and w, the Möbius representation of the dual fuzzy measure μ^d satisfies,*

$$\mu^d(A) = \begin{cases} \frac{1-w}{n-k-1}, & |A| = 1, \\ 0, & 1 < |A| \leq n - k - 1. \end{cases}$$

Proof. From the definition we can obtain that the dual ν^d of a k-interactive fuzzy measure will satisfy

$$\nu^d(A) = 1 - \left(w + \frac{|A'| - k - 1}{n - k - 1}(1 - w) \right) = (1-w)\frac{n - |A'|}{n - k - 1} = (1-w)\frac{|A|}{n - k - 1}$$

for $A, |A| \leq n - k - 1$. For $|A| = 1$, the Möbius values and standard representation values coincide and so we have,

$$\mu^d(A) = \frac{1 - w}{n - k - 1}, \ \forall \ |A| = 1.$$

For $1 < |A| \leq n-k-1$ we have

$$\begin{aligned}
\mu^d(A) &= \sum_{B\subseteq A} (-1)^{|A\setminus B|}\nu^d(B) \\
&= \sum_{B\subseteq A} (-1)^{|A\setminus B|}(1-w)\frac{|B|}{n-k-1} \\
&= \frac{1-w}{n-k-1}\sum_{B\subseteq A} (-1)^{|A\setminus B|}|B| \\
&= 0
\end{aligned}$$

□

Hence the key difference is that rather than all Möbius values under a given threshold being 0, the singletons are allocated a weight of $\frac{1-w}{n-k-1}$ (see Fig. 4).

However even though their duals are characterised by zero Möbus values for a range of subsets' cardinalities as per Proposition 2, the original k-interactive fuzzy measures do not share this property, and hence do not benefit from the reduction of variables in Möbius representation (although the fitting to data problem is indeed simplified, see [5]).

Remark 10. Note from the architecture that we also get k-intolerant and k-tolerant (corresponding with different k) fuzzy measures as a special case, when the outer WAM allocates 0 weight to the AM_2 input.

Excluding the singletons (which are fixed), the number of non-zero Möbius values for the dual of a k-intolerant fuzzy measure will be

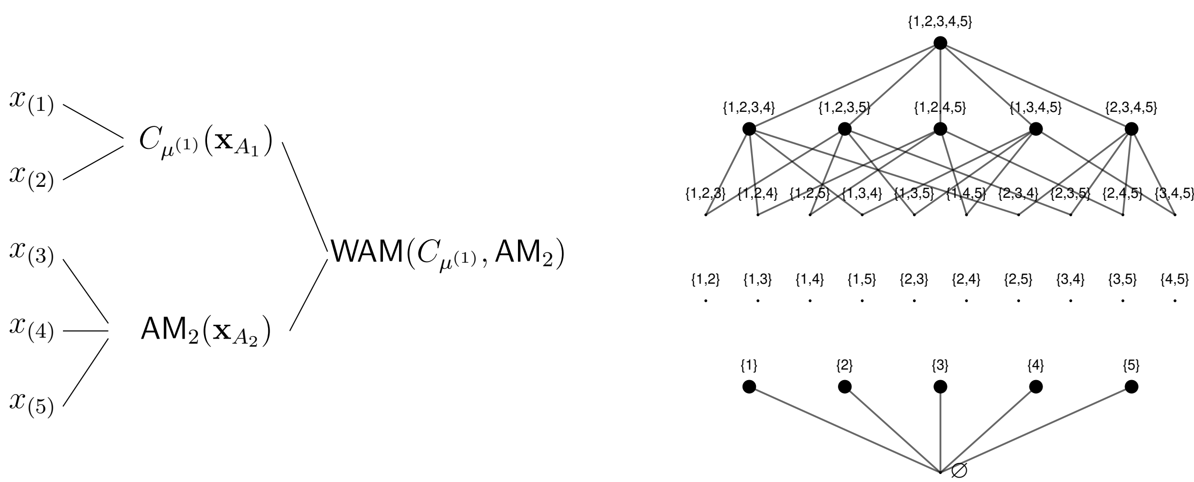
$$\sum_{i=n-k}^{n} \frac{n!}{i!(n-i)!} = \sum_{i=0}^{k} \frac{n!}{i!(n-i)!}.$$

For the constraints, again we need not worry about subsets of size up to $n-k-1$, and so consider only those of size $n-k$ or larger, giving us,

$$\sum_{i=n-k}^{n} i\frac{n!}{i!(n-i)!} = \sum_{i=0}^{k}(n-i)\frac{n!}{i!(n-i)!}$$

monotonicity constraints.

An advantage of such top-down interactive architectures is that monotonicity need only be considered across subsets of the larger sizes, e.g., for



(a) (b)

Figure 4: (a) Order-dependent architecture for the dual of a 1-interactive fuzzy measure, which averages the largest $n-k-1=3$ inputs but allows the smallest 2 inputs to interact, and (b) the corresponding interaction diagram showing Möbius values equal to zero for all $2 \leq |A| \leq 3$ (the singletons would all be equal to a fixed $\frac{1-w}{3}$).

$n = 8$ and considering subsets of size 6,7,8, we would require $8 + 7 \times 8 + 6 \times 28 = 232$ monotonicity constraints instead of 1016 required for general and for 3-additive fuzzy measures.

3.3. Generalising k-interactivity to k-lower and k-upper interactive fuzzy measures

A natural generalisation of the concepts of k-intolerance and k-interactivity is to allow singletons to carry non-equal weight. We propose the following definition.

Definition 11. A fuzzy measure is k-lower interactive for $k \in \{1, 2, \ldots, n-2\}$ when for all $A \subseteq N$ and $1 < |A| \leq n-k$, it holds that $\mu_A = 0$. The boundary cases of $k = 0$ and $k = n-1$ correspond with additive and general fuzzy measures respectively. A fuzzy measure is k-upper interactive if it is dual to a k-lower interactive fuzzy measure.

A k-lower interactive fuzzy measure takes a weighted mean of the highest $n-k$ inputs and allows for interaction in the lowest k inputs.

The architecture and interaction diagram representations of such fuzzy measures is identical to that of k-interactive fuzzy measures, except that the arithmetic mean is replaced with a weighted arithmetic mean in the first step. For example, in Fig. 4(a) we would have a 2-lower interactive fuzzy measure with AM_2 replaced with WAM_2.

The number of non-zero Möbius values we need to consider corresponds with k-intolerance and k-interactivity, except that we include the n singletons, hence a k-lower interactive fuzzy measure is defined by

$$n + \sum_{i=n-k+1}^{n} \frac{n!}{i!(n-i)!} = n + \sum_{i=0}^{k-1} \frac{n!}{i!(n-i)!}$$

variables for $0 < k < n-2$.

For the constraints, in addition to those corresponding with the larger subsets, we also will require $\mu_i \geq 0$ for each of the singletons. Hence, with respect to k, a k-lower interactive fuzzy measure will require consideration of

$$n + \sum_{i=n-k+1}^{n} i\frac{n!}{i!(n-i)!} = n + \sum_{i=0}^{k-1} (n-i)\frac{n!}{i!(n-i)!}$$

constraints, with $0 < k < n-2$.

Table 2 gives an idea of the savings that can be achieved in the fitting problem by using the k-lower interactive fuzzy measure simplification for $n = 5$, $n = 10$ and $n = 100$.

Remark 12. While we do not explore it in more detail here, in this framework one could also consider fixing values for sets of up to cardinality $|A| = n-k$, resulting in ordered weighted aggregation of the largest $n-k$ inputs. This would limit the variables and constraints required in a similar way and achieve similar cost reduction.

4. Two-step Choquet integrals with WAM-equivalent aggregation at the second step

We now consider hierarchical aggregation structures based on partitions or coverings of the input set, where the first aggregation step is performed

Table 2: Unidentified parameters (var) and constraints (constr) required for k-lower interactive fuzzy measures

	$n=5$		$n=10$		$n=100$	
k	var	constr	var	constr	var	constr
1	6	10	11	20	101	200
2	11	30	21	110	201	10100
3	21	60	66	470	5151	495200
4	31	80	186	1310	166851	16180100
9			1023	5120	$2.03 \cdot 10^{11}$	$1.87 \cdot 10^{13}$

using a set of Choquet integrals and these values are then aggregated at the second step using a weighted arithmetic mean (or additive Choquet integral). This framework is often referred to as a two-step Choquet integral. We will use the notation $\mathbf{x}_A$ to denote a restriction of the vector $\mathbf{x}$ to its components corresponding to the elements in A, e.g., for $\mathbf{x} = (0.2, 0.7, 0.3, 0.5)$ and $A = \{2, 3\}$ then $\mathbf{x}_A = (0.7, 0.3)$.

Some of the results below will make use of the fact that any fuzzy measure of k variables can be embedded into a fuzzy measure of $n > k$ variables, where the variables $k+1, k+2, \ldots, n$ are allocated zero weight. In Möbius representation, all supersets of these elements k will be zero. An example for $k = 3$ and $n = 5$ is shown in Fig. 5. Embeddings of fuzzy measures in this way can also be viewed in terms of *dummy criteria* and *dummy coalitions* (see, e.g., [7–9, 15]).

From this and the result that any weighted arithmetic mean of fuzzy measures is also a fuzzy measure we can articulate the following proposition.

Proposition 3. *Consider an input vector $\mathbf{x} = (x_1, \ldots, x_n)$ and a covering $\{A_1, \ldots, A_m\}$, i.e., one for which $A_i \neq \emptyset$ for all i and $\bigcup_{i=1}^m A_i = N$. For any weighted arithmetic mean of Choquet integrals $C_{\mu^{(1)}}(\mathbf{x}_{A_1})$, $C_{\mu^{(2)}}(\mathbf{x}_{A_2})$, $\ldots, C_{\mu^{(m)}}(\mathbf{x}_{A_m})$ there exists an equivalent single Choquet integral $C_\mu(\mathbf{x})$.*

Proof. The result follows from the fact that, using Möbius representation, each of the integrals $C_{\mu^{(i)}}(\mathbf{x}_{A_i})$ can be expressed as embedded fuzzy measures with respect to the n-variate case and hence the final fuzzy measures weights can be obtained from a linear combination of these m fuzzy measures. Each of the Möbius weights in the overall fuzzy measure μ is then simply the weighted arithmetic mean of corresponding coefficients. □

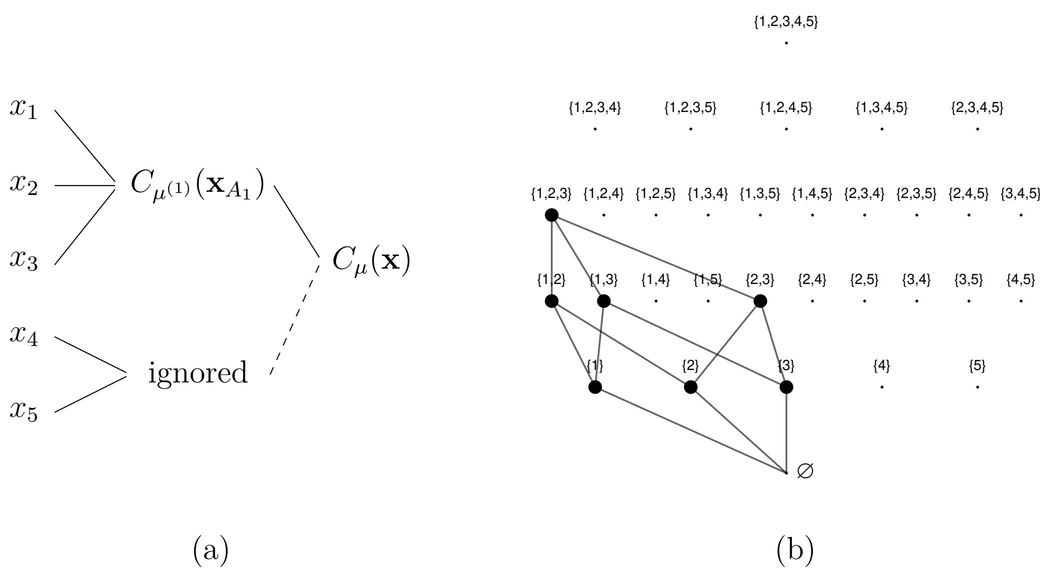


Figure 5: A 3-variate fuzzy measure defined in reference to the inputs x_1, x_2, x_3 embedded into a 5-variate fuzzy measure with these inputs mapped as they are and x_4, x_5 are allocated zero weight. The architecture is shown in (a) while the interaction diagram corresponding with the 5-variate fuzzy measure is shown in (b).

Remark 13. See previous works [20, 25] for alternative approaches to the proof.

While this may seem straight forward for partitions, the following example shows how it works when there is overlap between the subsets.

Example 2. Suppose

$$\begin{aligned} f &= \mathsf{WAM}(C_{\mu^{(1)}}(x_1, x_2, x_3), C_{\mu^{(2)}}(x_3, x_4, x_5)) \\ &= w_1 \cdot C_{\mu^{(1)}}(x_1, x_2, x_3) + w_2 \cdot C_{\mu^{(2)}}(x_3, x_4, x_5) \end{aligned}$$

In this case, $A_1 = \{1, 2, 3\}, A_2 = \{3, 4, 5\}$. In Möbius representation, we have

$$C_{\mu^{(1)}}(\mathbf{x}_{A_1}) = \mu^{(1)}_{\{1\}} x_1 + \mu^{(1)}_{\{2\}} x_2 + \mu^{(1)}_{\{3\}} x_3 + \mu^{(1)}_{\{1,2\}} \min(x_1, x_2) + \cdots$$
$$\cdots + \mu^{(1)}_{\{1,2,3\}} \min(x_1, x_2, x_3)$$

and

$$C_{\mu^{(2)}}(\mathbf{x}_{A_2}) = \mu^{(2)}_{\{3\}}x_3 + \mu^{(2)}_{\{4\}}x_4 + \mu^{(2)}_{\{5\}}x_5 + \mu^{(2)}_{\{3,4\}}\min(x_3,x_4) + \cdots$$

$$\cdots + \mu^{(2)}_{\{3,4,5\}}\min(x_3,x_4,x_5).$$

The overall fuzzy measure such that $f = C_\mu(\mathbf{x}) = f(\mathbf{x})$ can be obtained by aggregating $\mu^{(1)}$ and $\mu^{(2)}$ according to the weights w_1, w_2. We have

$$\begin{array}{llll} \mu_{\{1\}} = w_1\mu^{(1)}_{\{1\}} & \mu_{\{2\}} = w_1\mu^{(1)}_{\{2\}} & \mu_{\{3\}} = w_1\mu^{(1)}_{\{3\}} + w_2\mu^{(2)}_{\{3\}} & \mu_{\{4\}} = w_2\mu^{(2)}_{\{4\}} \\ \quad\cdots & \quad\cdots & \mu_{\{1,2,3\}} = w_1\mu^{(1)}_{\{1,2,3\}} & \mu_{\{3,4,5\}} = w_2\mu^{(2)}_{\{3,4,5\}} \end{array}$$

and zero for all subsets not involved in the calculation of either C_1 or C_2. Hence, although it is a fuzzy measure defined across the 5 variables and 31 subsets, it requires vastly fewer of the weights to be determined.

Figure 6 shows (a) the hierarchical structure and (b) an interaction diagram indicating the fuzzy measure weights considered (vertices without edges leading into them from below would have a zero value). For comparison with a partition, (c) and (d) show the respective architecture and interaction diagram with $A_1 = \{1,2,3\}, A_2 = \{4,5\}$.

We can obtain a number of propositions and corollaries from this idea.

Proposition 4. *Consider any number of k_i-additive fuzzy measures defined over the subsets of a covering $\bigcup_{i=1}^m A_i = N$ where the level of additivity k_i may differ for each subset. Then, the overall fuzzy measure corresponding with a Choquet integral C_μ equivalent to taking a weighted mean of the component Choquet integrals $C_{\mu^{(i)}}$ is at most k-additive, where $k = \max k_i$.*

Proof. Each of the fuzzy measures $\mu^{(i)}$ can be embedded into an n-variate fuzzy measure where all subsets larger than k_i will be allocated zero weight in Möbius representation. The overall fuzzy measure's components are calculated as a weighted mean of the values attached to each of the subsets, and hence any subsets of size larger than $k = \max k_i$ will be determined by aggregating zeros. □

Corollary 14. *A weighted mean of k-additive n-variate Choquet integrals will be equivalent to a Choquet integral with respect to a fuzzy measure that is at most k-additive.*

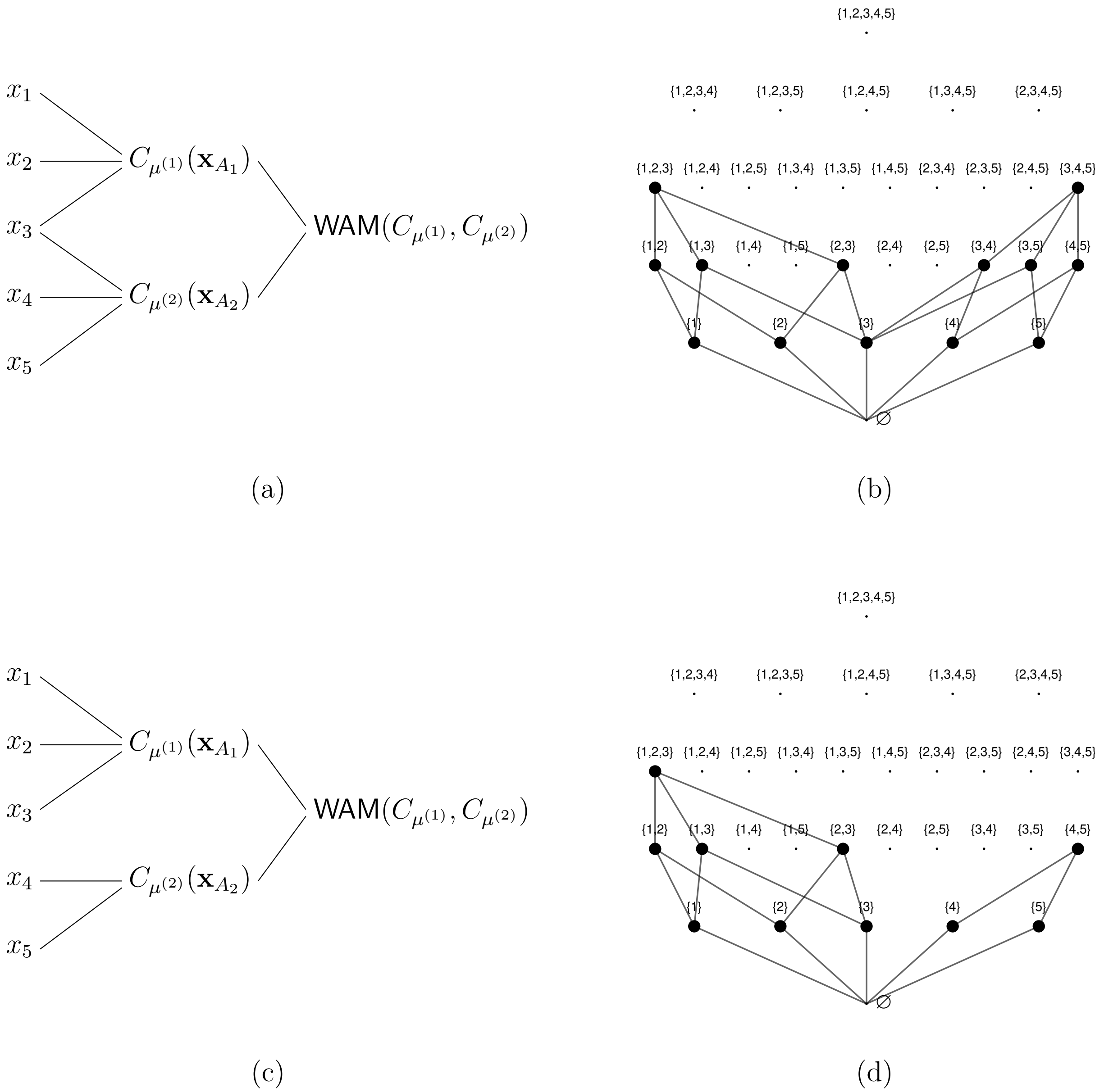


Figure 6: Hierarchical fuzzy measure architectures involving (a) a covering and (c) a partition, beside interaction diagram visualisations of the resulting fuzzy measures in (b) and (d). Larger nodes and edges from below indicate non-zero Möbius values.

Proposition 5. *Consider any number of k_i-lower interactive fuzzy measures defined over the subsets of a covering $\bigcup_{i=1}^{m} A_i = N$ where the level of lower interactivity k_i may differ for each subset. Then, the overall fuzzy measure corresponding with a Choquet integral C_μ equivalent to taking a weighted mean of the component Choquet integrals $C_{\mu^{(i)}}$ is at most k-lower interactive, where $k = n - \min(|A_i| - k_i)$.*

Proof. A k-lower interactive fuzzy measure satisfies $\mu_A = 0$ for $1 < |A| \leq n-k$, which means that for the component fuzzy measures, all subsets of size $2, 3, \ldots, |A_i| - k_i$ will be zero-valued in Möbius representation. In the overall fuzzy measure, all subsets of size $2, \ldots, \min(|A_i| - k_i)$ will be calculated as an aggregation of zeros and hence the overall fuzzy measure will be $n - \min(|A_i| - k_i) = k$-lower interactive. □

Corollary 15. *A weighted mean of k-lower interactive n-variate fuzzy measures will be at most k-lower interactive.*

Since k-intolerant fuzzy measures are a special case of k-lower interactive fuzzy measures, we also get the following corollary.

Corollary 16. *A weighted mean of k-intolerant fuzzy measures will be at most k-intolerant.*

Similar results pertaining to the dual concepts can be determined as consequences of the above. The following establishes that any dual of a hierarchical-architecture based fuzzy measure limits the interactions to the same subsets.

Proposition 6. *Consider a fuzzy measure corresponding with a two-step Choquet integral and where the second step of aggregation is equivalent to a WAM with respect to the weighting vector $\mathbf{w}$. Then, the dual of this function will also be a two-step Choquet integral with the second step equivalent to a WAM with respect to the same $\mathbf{w}$.*

Proof. This follows from the Möbius values dual calculation in Eq. (5). Each subset's value only takes into account its supersets and only subsets $B \subseteq A_i$ for some i are non-zero valued. The values for each subset of a particular A_i are hence calculated independently of the other A_i coalitions in the partition. Further, since the values within each A_i correspond with an embedded fuzzy measure multiplied by w_i, the dual calculation will also result in Möbius values summing to w_i. □

While the number of parameters required for general fuzzy measures is $2^n - 1$, fuzzy measures corresponding with two-step Choquet integrals, where the second step of aggregation is equivalent to a weighted arithmetic mean, have far fewer variables than the general fuzzy measure.

Given an input set $\mathbf{x} \in [0,1]^n$ and a partition $\bigcup_{i=1}^{m} A_i = N$, a fuzzy measure corresponding with the weighted mean of each $C_{\mu^{(i)}}(\mathbf{x}_{A_i})$ will have, at most

$$-m + \sum_{i=1}^{m} 2^{|A_i|} \tag{7}$$

non-zero weights.

This will also be the upper bound for coverings, where some variables will be repeated within the non-disjoint A_i subsets.

Example 2 above helps illustrate this. It is worth noting that the savings go far beyond even what is achieved by k-additive fuzzy measures and the order-dependent architectures. In the case of k-additive fuzzy measures, all subsets are considered to interact. For $n = 16$ and $k = 4$ we would still be considering 2516 unknowns. If, on the other hand, we know that there are 4 distinct groups of 4 interacting variables, we can reduce the number of unknown variables to 60.

However, it is not only the number of variables reduced but also the number of monotonicity constraints. Unlike the k-additive fuzzy measures, we need only consider monotonicity relationships for each of the subgroups.

Given an input set $\mathbf{x} \in [0,1]^n$ and a partition $\{A_1, \dots, A_m\}$ of N, a fuzzy measure corresponding with the weighted mean of each $C_{\mu^{(i)}}(\mathbf{x}_{A_i})$ needs only consider the constraints

$$\sum_{B \subseteq S : j \in B} \mu_B \;\geq\; 0, \text{ for each } j \in S, \text{ for all } S \subseteq A_i, i = 1, \dots, m, \tag{8}$$

of which there will be

$$\sum_{i=1}^{m} \sum_{j=1}^{|A_i|} j \frac{|A_i|!}{j!(|A_i| - j)!} = \sum_{i=1}^{m} |A_i| 2^{|A_i|-1}, \tag{9}$$

which is a reduction from $O(n2^{n-1})$ to $O(n2^{k-1})$, where $k = \max |A_i|$ and $n = \sum |A_i|$. This owes to supersets of each A_i being fixed with zero weight in Möbius representation.

Table 3: Data, non-negativity, monotonicity, and boundary constraints (in Möbius form, excluding the introduced residual coefficients) for $n = 4$ where $\{1,2\}$ and $\{3,4\}$ constitute independent subgroups.

$\mu(\{1\})$	$\mu(\{2\})$	$\mu(\{1,2\})$	$\mu(\{3\})$	$\mu(\{4\})$	$\mu(\{3,4\})$		RHS
$x_1^{(1)}$	$x_2^{(1)}$	$\min(x_1^{(1)}, x_2^{(1)})$	$x_3^{(1)}$	$x_4^{(1)}$	$\min(x_3^{(1)}, x_4^{(1)})$	$=$	$y^{(1)}$
$\dots$		$\dots$			$\dots$	$\dots$	$\dots$
$x_1^{(d)}$	$x_2^{(d)}$	$\min(x_1^{(d)}, x_2^{(d)})$	$x_3^{(d)}$	$x_4^{(d)}$	$\min(x_3^{(d)}, x_4^{(d)})$	$=$	$y^{(d)}$
1						$\geq$	0
	1					$\geq$	0
			1			$\geq$	0
				1		$\geq$	0
1		1				$\geq$	0
	1	1				$\geq$	0
			1		1	$\geq$	0
				1	1	$\geq$	0
1	1	1	1	1	1	$=$	1

Example 3. Suppose we have

$$C_\mu(\mathbf{x}) = \mathsf{WAM}(C_{\mu^{(1)}}(x_1, x_2), C_{\mu^{(1)}}(x_3, x_4)).$$

In this case, $n = 4$, $|A_1| = |\{1,2\}| = |A_2| = |\{3,4\}| = 2$. The calculations for number of variables and monotonicity constraints yield 6 and 4 respectively. The constraints required, along with the data constraints (not including the residual coefficients) are shown in Table 3.

For overlapping sub-groups, i.e., where we have a covering rather than a partition, the calculations will differ somewhat. While there will be some redundancy in the variables considered, we will require additional constraints. For example, assuming an element j belongs to, at most, two subsets A_i and A_k, in addition to the constraints for a partition included in (8), we would need to also consider

$$\sum_{B \subseteq S: j \in B} \mu_B \;\; \geq \;\; 0, \text{ for all } S \subseteq A_i \cup A_k \text{ such that } j \in S, S \nsubseteq A_i, A_j. \quad (10)$$

Then clearly if j belonged to more than 2 subsets we would need to consider subsets including members from each of the non-disjoint sets along with j.

Example 4. Suppose we have the two-step Choquet integral as depicted in Fig.6(a) and (b), i.e.,

$$C_\mu(\mathbf{x}) = \mathsf{WAM}(C_{\mu^{(1)}}(x_1, x_2, x_3), C_{\mu^{(2)}}(x_3, x_4, x_5)).$$

In this case, $n = 5$, $|A_1| = |A_2| = 3$. Here we have 7 variables each corresponding to all subsets of A_1 and A_2, however $\mu_{\{3\}}$ is repeated and so we only have 13 altogether. For monotonicity we will require 2 monotonicity constraints for each of the pairs within the coalitions, i.e., for

$$\{1, 2\}, \{1, 3\}, \{2, 3\}, \{3, 4\}, \{3, 5\}, \{4, 5\},$$

and 3 monotonicity constraints for each of $\{1, 2, 3\}$ and $\{3, 4, 5\}$. However, we also need one constraint for each of the supersets of $\{3\}$ that include elements from both coalitions, i.e., for

$$\{1, 3, 4\}, \{1, 3, 5\}, \{2, 3, 4\}, \{2, 3, 5\},$$

$$\{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 3, 4, 5\}, \{2, 3, 4, 5\},$$

$$\text{and } \{1, 2, 3, 4, 5\}$$

For instance, in the case of $\{1, 3, 4\}$ we need

$$\mu_{\{1,3,4\}} + \mu_{\{3,4\}} + \mu_{\{1,3\}} + \mu_{\{3\}} \geq 0,$$

which simplifies to

$$\mu_{\{3,4\}} + \mu_{\{1,3\}} + \mu_{\{3\}} \geq 0,$$

which is not taken into account by the monotonicity constraints corresponding with subsets of A_1 or A_2. We can reiterate that for the additional subsets, we only need to consider the constraints based on the inclusion of $\{3\}$, i.e., the constraint based on inclusion of $\{1\}$ would simplify to $\mu_{\{1,3\}} + \mu_{\{1\}} \geq 0$, which is redundant with those already considered. This gives 32 constraints in total, whereas usually for a 5-variate fuzzy measure we need 80.

Table 4 is provided to give an idea of the reduction in variables and constraints that can be achieved using hierarchical fuzzy measures based on partitions.

It is evident that assuming the data follows a hierarchical structure with a weighted arithmetic mean aggregation at the outer layer is able to achieve a significant reduction in variables and constraints, especially if the variables

Table 4: Unidentified parameters (var) and constraints (constr) required for hierarchical fuzzy measures based on partitions with disjoint subsets of at most cardinality k

	$n=5$		$n=10$		$n=100$	
k	var	constr	var	constr	var	constr
1	5	5	10	10	100	100
2	7	9	15	20	150	200
3	10	16	22	37	232	397
4	16	33	33	68	375	800
5	31	80	62	160	620	1600
10			1023	5120	10230	51200

can be grouped into a partition. The overall fuzzy measures may not be as flexible as k-additive fuzzy measures, where interaction is modelled between all variables, however interaction in groups of 2, 3 or more can still be considered. Limiting interaction to particular groups may also be realistic and practical in a number of contexts to avoid incorporating interaction effects in the data that only appear to be present by chance.

5. Two-step Choquet integral with interaction at the second step

We have observed that when the second step of aggregation is performed with respect to an additive fuzzy measure (WAM), the overall aggregation can be expressed as a single fuzzy measure. On the other hand, when the outer aggregation is non-additive, it will not always be the case that the hierarchy collapses to a single aggregation [25]. Here is a counterexample.

Example 5. Suppose we have the hierarchical structure,

$$f(x_1,x_2,x_3,x_4,x_5) = C_{\mu^*}(\mathsf{WAM}_{\mathbf{w}^{(1)}}(x_1,x_2), \mathsf{WAM}_{\mathbf{w}^{(2)}}(x_3,x_4,x_5))$$

Let $\mathbf{w}^{(1)} = (0.2, 0.8), \mathbf{w}^{(2)} = (0.1, 0.7, 0.2)$ and for the fuzzy measure let

$$\mu^*_{\{1\}} = 0.9, \mu^*_{\{2\}} = 0.4, \mu^*_{\{1,2\}} = -0.3.$$

A Choquet integral is comonotone additive, however note for the comonotone input vectors

$$\mathbf{x} = (0.9, 0.4, 0.2, 0.5, 0.7), \mathbf{y} = (0.9, 0.7, 0.2, 0.75, 0.8),$$

we have

$$f(\mathbf{x}) = 0.504, \qquad f(\mathbf{y}) = 0.7365,$$
$$f(\mathbf{x}+\mathbf{y}) = 1.2375,$$

and hence $f(\mathbf{x}) + f(\mathbf{y}) = 1.2405 \neq f(\mathbf{x}+\mathbf{y})$.

Nonetheless such structures will inherit some salient properties of the Choquet integrals, see also [20].

Proposition 7. *A two-step hierarchical aggregation structure with Choquet integrals at the first and second step will result in a function that is homogeneous, shift-invariant, monotone, and piecewise-linear.*

Proof. The Choquet integral can be expressed as a linear combination of piecewise linear functions and hence homogeneity, shift-invariance, monotonicity, and piecewise-linearity will be preserved. □

We now consider two examples of fuzzy measures with special characteristics in Möbius representation and how these correspond with or approximate hierarchical aggregation methods.

5.1. Nonadditive between-group interactions of minimum/maximum aggregated coalitions

One instance where we can use the Choquet integral in the second step of aggregation and arrive at a structure equivalent to a single Choquet integral is where the aggregators at the first step are all minimum functions.

Proposition 8. *For a two-step Choquet integral, if the fuzzy measure at the second step μ^* is non-additive and the fuzzy measures $\mu^{(i)}, i = 1, \ldots, m$ at the first step each model the minimum, then the overall aggregation will correspond with a single Choquet integral with respect to a fuzzy measure μ.*

Proof. Let $M = \{1, \ldots, m\}$. From the expression of the Choquet integral in Möbius representation (Eq. (1)) we have,

$$C_{\mu^*}(\min(\mathbf{x}_{A_1}), \ldots, \min(\mathbf{x}_{A_m})) = \sum_{B \subseteq M} \mu^*_B \min_{i \in B}(\min_{j \in A_i} x_j) = \sum_{B \subseteq M} \mu^*_B \min_{j \in \{\bigcup\limits_{i \in B} A_i\}} x_j.$$

Since all sets $\{\bigcup\limits_{i \in B} A_i\}$ coincide with one of the subsets of N, there will exist an equivalent overall fuzzy measure μ defined on N, with all Möbius values zero except for those corresponding with all unions of the A_i. □

The upshot of such structures is that we can transform the dataset according to the A_i coalitions and fit to μ^*, however still gain an overall understanding of the aggregation process by interpreting μ. The following example considers the case of a partition.

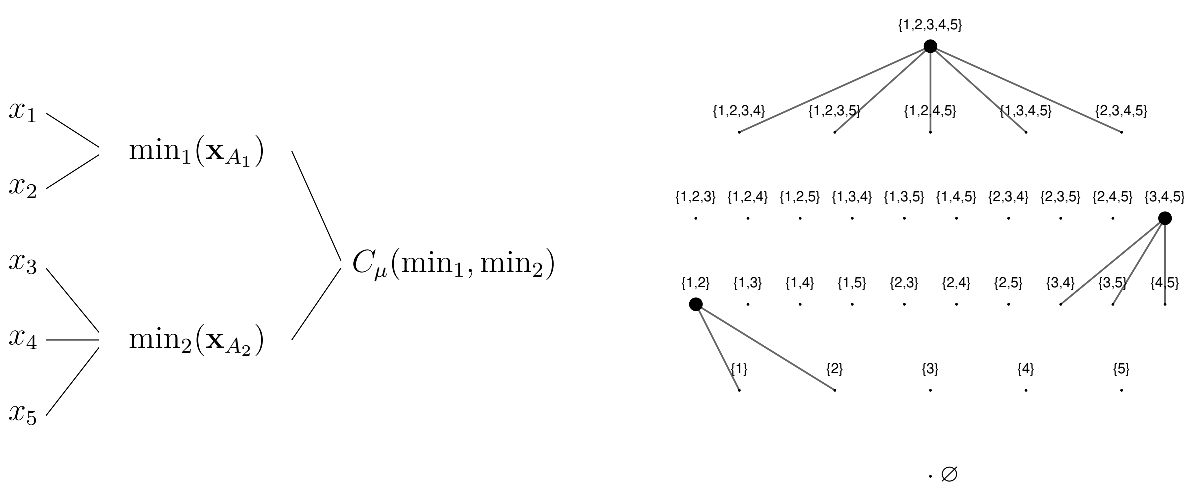
Example 6. Let the function architecture be such that

$$f = C_{\mu^*}(\min(x_1, x_2), \min(x_3, x_4, x_5)).$$

Using Möbius representation, we will have

$$\begin{aligned} f &= \mu_1^* \min(x_1, x_2) + \mu_2^* \min(x_3, x_4, x_5) \\ &\quad + \mu_{12}^* \min(\min(x_1, x_2),\ \min(x_3, x_4, x_5)) \\ &= \mu_1^* \min(x_1, x_2) + \mu_2^* \min(x_3, x_4, x_5) + \mu_{12}^* \min(x_1, x_2, x_3, x_4, x_5) \end{aligned}$$

Hence this is equivalent to a single Choquet integral defined for $n = 5$ with $\mu_{12} = \mu_1^*, \mu_{345} = \mu_2^*, \mu_{12345} = \mu_{12}^*$ and all other Möbius values equal to 0. Fig. 7 depicts the hierarchical structure and the interaction diagram.



(a) (b)

Figure 7: (a) Hierarchy taking a Choquet integral of minimum functions and (b) the interaction diagram of the corresponding overall fuzzy measure.

Remark 17. Note in 7(b) only the three subsets are non-zero, with all edges shown to each of the subsets of one less cardinality.

Such overall fuzzy measures can be understood in terms of what are sometimes referred to as partnerships in the study of non-additive games [9], where in standard representation the coalitions in the partition A_i satisfy $\nu(S \cup T) = \nu(S)$ for all $T \subsetneq A_i$ and all $S \subset N \setminus A_i$. Partnerships are interpreted as behaving as a single hypothetical player in a reduced game.

We have a similar result when the first step aggregations are performed using the maximum.

Proposition 9. *For a two-step Choquet integral, if the fuzzy measure at the second step μ^* is non-additive and the fuzzy measures $\mu^{(i)}, i = 1, \ldots, m$ at the first step each model the maximum, then the overall aggregation will correspond with a single Choquet integral with respect to a fuzzy measure μ.*

Proof. The Choquet integral can be expressed as a linear combination of maximum functions using Möbius representation and considering the relationship with the dual aggregation function, i.e.

$$C_{\mu^d}(\mathbf{x}) = 1 - \sum_{A \subseteq N} \mu_A \min_{i \in A}(1 - x_i) = 1 - \sum_{A \subseteq N} \mu_A (1 - \max_{i \in A} x_i) = \sum_{A \subseteq N} \mu_A \max_{i \in A} x_i.$$

Hence, following the same logic as in the proof of Proposition 8, there will exist an equivalent overall fuzzy measure μ defined on N. □

Propositions 8 and 9 are also established in [20, 25]. We can then turn to the case of two-step Choquet integrals where both minimum and maximum functions are involved in the first step. For general coverings, the function will not necessarily reduce to a single Choquet integral. For example, we could not model

$$f = C_{\mu^*}(\min(x_1, x_2), \max(x_2, x_3)),$$

because according to the minimum-aggregated group we require $\mu_{\{2\}} = 0$ and according to the maximum-aggregated group we require $\mu_{\{2\}} > 0$ and so we have a contradiction.

However, it will always be possible when the aggregations at the first step are defined over a partition, see also [20].

Proposition 10. *For a two-step Choquet integral, if the fuzzy measure at the second step μ^* is non-additive and the fuzzy measures $\mu^{(i)}, i = 1, \ldots, m$ at the* *first step defined over a partition* $\bigcup\limits_{i\in M} A_i = N$ *all either model a minimum or maximum, then the overall aggregation will correspond with a single Choquet integral with respect to a fuzzy measure* μ.

Proof. We need only show that the arguments in the second step of integration $\min(C_{\mu^{(1)}}, C_{\mu^{(2)}})$, $\min(C_{\mu^{(1)}}, C_{\mu^{(3)}})$, etc., can be expressed as Choquet integrals over the unions of their component sets. Since for functions of the form

$$\min\left(\min_{j\in A_1} x_j, \min_{j\in A_2} x_j, \ldots, \max_{j\in A_{m-1}} x_j, \max_{j\in A_m} x_j\right),$$

any of the minimum functions can be expressed as a composition of maximums, i.e. of each of the singletons,

$$\min\left(\max_{j\in\{1\}} x_j, \max_{j\in\{2\}} x_j, \ldots, \max_{j\in A_{m-1}} x_j, \max_{j\in A_m} x_j\right),$$

it follows from Proposition 9 that the resulting expressions will also be reducible to a single Choquet integral. □

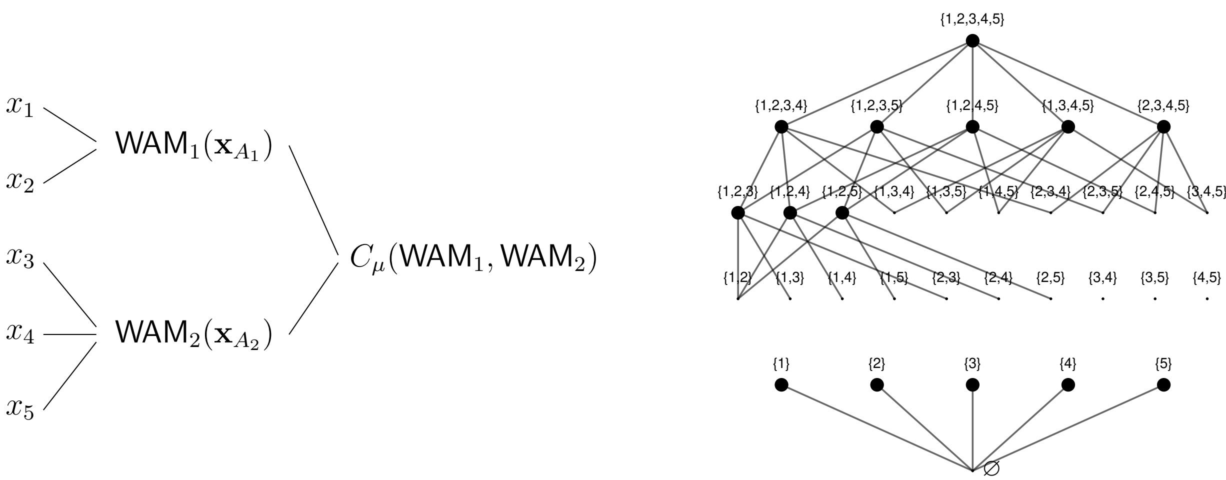
While such functions are equivalent to a single Choquet integral, considering the learning problem in a hierarchical framework reduces the weight identification problem to consideration of μ^* only, with the fuzzy measures $\mu^{(i)}$ defining the Choquet integrals at the first step fixed. We consider the covering or partition $\bigcup\limits_{i\in M} A_i = N$. Rather than transforming the input set so that inputs are mapped to $\min(\mathbf{x}_A)$ for all $A \subseteq N$, we instead map the inputs to each $B \subseteq M$ with the variable μ^*_B corresponding to the transformed inputs based on the aggregation at the first step of integration. The monotonicity and boundary constraints need only be considered as they pertain to $B \subseteq M$.

After fitting, if the first step of aggregation is based on a partition, the behaviour across the overall fuzzy measure can be interpreted by mapping the μ^*_B values to each μ_A with $A = \bigcup\limits_{i\in B} A_i$.

Remark 18. It is worth noting from the above that, although it would be possible to fit to the reduced data mapping corresponding with μ^* in the case of a covering, we would not necessarily be able to extract the values for interpretation of the overall fuzzy measure, since the unions of A_i would not necessarily be unique, i.e., in Example 6 if we also had a third argument $\min(x_1, x_2, x_4)$, then μ^*_{12}, μ^*_{23} and μ^*_{123} will all be associated with $\min(x_1, x_2, x_3, x_4, x_5)$.

5.2. *Fuzzy measures that approximate within- and between-coalition interaction*

When Möbius values are considered a proxy for interaction, we can make arbitrary choices as to which variables interact. One way of considering subsets and interactions is in terms of between-coalition interactions, i.e., similar to partnerships, we consider subsets more or less behaving like a single entity when it comes to interaction, with additive behaviour within the group and only allowing interaction effects with other groups as a whole. In this case, only supersets of the coalitions along with all the singletons would be allocated a value in Möbius representation.



(a) (b)

Figure 8: (a) Hierarchy taking a Choquet integral of weighted arithmetic means and (b) interaction diagram for an approximating overall fuzzy measure whereby only the singletons and supersets of $\{1,2\}$ and $\{3,4,5\}$ are not fixed at 0.

Example 7. With $A_1 = \{1,2\}$ and $A_2 = \{3,4,5\}$, we can allow non-zero Möbius values corresponding only with the singletons and supersets of either A_1 or A_2. We don't consider interaction between the first and third element ($\mu_{\{1,3\}} = 0$), however we do consider interaction between $\{1,2\}$ and $\{3\}$. We note from the interaction diagram in Fig. 8(b) that this could be considered a special case of the k-lower interactive fuzzy measures, however it also can be interpreted, at least in semantics, as an approximation of the hierarchical architecture shown in Fig. 8(a).

This approach can hence also result in a reduction of variables and constraints. From the observation regarding the relationship to k-lower and k-upper interactive fuzzy measures we can surmise that similar reductions in variables and constraints could be achieved depending on the coalition cardinalities.

6. Discussion and conclusions

We have considered hierarchical fuzzy measures and the two-step Choquet integral toward the goal of learning nonadditive models from data and user preferences. In particular, we focused on Möbius representation and certain characteristics that have an effect on the number of unknown parameters, constraints, and setting up of the objective functions. These features serve as simplifications that allow aggregation by the Choquet integral to be implemented in a much broader range of scenarios, with the learning problem becoming tractable even for large dimensionality. The proposed models effectively explore sparsity of the matrices of constraints and reduction of the model parameters. While such models do not exhibit the full flexibility of a Choquet integral with respect to a general fuzzy measure with $2^n - 1$ parameters, a trade-off is accomplished that still allows utilisation of its unique ability to incorporate explainable interaction effects. It is worth noting that such models may be able to closely approximate a broad range of practical situations, however our considerations have not been exhaustive. For example, the p-symmetric fuzzy measures introduced in [21], which allow sets of indifference, could also be used (as well as many other simplifications) to define a reduced set of constraints and variables.

Of high value to practitioners are the results pertaining to the use of Choquet integrals where the input set is partitioned into interacting coalitions, with such models achieving a drastic reduction in the number of variables and constraints. We also proposed a simplification referred to as k-lower (and k-upper) fuzzy measures, which, similar to the k-additive fuzzy measures allow reduction in the number of variables used but also in the number of monotonicity constraints required to learn values from data. This simplification generalises k-intolerant and k-interactive fuzzy measures, allowing interaction to be modelled for the lower or higher inputs and additive weighted aggregation to be applied to the remaining inputs.

As we have focused our results toward Möbius representation, it is worth making mention of the fact that interaction indices can also be calculated efficiently from these values. For instance, Shapley values $\phi(i)$ and interaction indices $I(A)$ from Möbius values are calculated as [11]

$$\phi(i) = \sum_{B:i\in B} \frac{1}{|B|}\mu_B \tag{11}$$

$$I(A) = \sum_{B:A\subseteq B} \frac{1}{|B|-|A|+1}\mu_B, \tag{12}$$

wheras the non-modularity index [26] is calculated as

$$d_\mu(A) = \sum_{B\subseteq A,|B|\geq 2} \frac{|B|}{|A|}\mu_B. \tag{13}$$

While the most significant implication of these collected results is the ability to use fuzzy measures and integrals for larger numbers of inputs, the perspective we have gained here on interaction and sparsity could also be capitalised on toward regularisation and other statistics-based methods.

Acknowledgements

The work was supported by the Australian Research Council Discovery project DP210100227.

References

[1] S. Angilella, S. Greco, and B. Matarazzo. Non-additive robust ordinal regression: a multiple criteria decision model based on the Choquet integral. *Eur. J. Operat. Res.*, 201:277 – 288, 2010.

[2] G. Beliakov, H. Bustince, and T. Calvo. *A Practical Guide to Averaging Functions.* Springer, New York, 2016.

[3] G. Beliakov and S. James. Citation based journal ranks: The use of fuzzy measures. *Fuzzy Sets and Systems*, 167:101–119, 2011.

[4] G. Beliakov, S. James, and J-Z. Wu. *Discrete Fuzzy Measures: Computational Aspects.* Springer, Berlin, Heidelberg, 2019.

[5] G. Beliakov and J.-Z. Wu. Learning fuzzy measures from data: Simplifications and optimisation strategies. *Information Sciences*, 494:100 – 113, 2019.

[6] G. Choquet. Theory of capacities. *Ann. Inst. Fourier*, 5:1953–1954, 1953.

[7] K. Fujimoto. Cardinal-probabilistic interaction indices and their applications: A survey. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, 7(2):79 – 85, 2003.

[8] K. Fujimoto. Cooperative game as non-additive measure. In *Non-Additive Measures, Studies in Fuzziness and Soft Computing 310*, pages 131 – 171. Springer International Publishing, Switzerland, 2014.

[9] K. Fujimoto, I. Kojadinovic, and J.-L. Marichal. Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. *Games and Economic Behavior*, 55:72 – 99, 2006.

[10] M. Grabisch. The applications of fuzzy integrals in multicriteria decision making. *Europ. J. Operations Research*, 89:445–456, 1996.

[11] M. Grabisch. k-order additive discrete fuzzy measures and their representation. *Fuzzy sets and systems*, 92(2):167–189, 1997.

[12] M. Grabisch. *Set Functions, Games and Capacities in Decision Making.* Springer, Berlin, New York, 2016.

[13] 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 J. of Operational Research*, 186(2):766–785, 2008.

[14] M. Grabisch, J.-L. Marichal, and M. Roubens. Equivalent representations of set functions. *Mathematics of Operations Research*, 25(2):157–178, 2000.

[15] M. Grabisch and M. Roubens. An axiomatic approach to the concept of interaction among players in cooperative games. *Int. J. Game Theory*, 28:547 – 565, 1999.

[16] A. Honda and S. James. Parameter learning and applications of the inclusion-exclusion integral for data fusion and analysis. *Information Fusion*, 56:28–38, 2020.

[17] J.-L. Marichal. k-intolerant capacities and Choquet integrals. *European Journal of Operational Research*, 177:1453–1468, 2007.

[18] A. Mendez-Vazquez, P. Gader, J.M. Keller, and K. Chamberlin. Minimum classification error training for Choquet integrals with applications to landmine detection. *IEEE Trans. Fuzzy Syst.*, 16:225 – 238, 2008.

[19] R. Mesiar. Generalizations of k–order additive discrete fuzzy measures. *Fuzzy Sets and Systems*, 102:423–428, 1999.

[20] R. Mesiar and D. Vivona. Two-step integral with respect to fuzzy measure. *Tatra Mountains Mathematical Publications*, 16:359–368, 1999.

[21] P. Miranda, M. Grabisch, and P. Gil. p-symmetric fuzzy measure. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(supp01):105–123, 2002.

[22] T. Murofushi, K. Fujimoto, and M. Sugeno. Canonical separated hierarchical decomposition of the Choquet integral over a finite set. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 6(3):257 – 272, 1998.

[23] B. J. Murray, M. A. Islam, A. J. Pinar, D. T. Anderson, G. J. Scott, T. C. Havens, and J. M. Keller. Explainable AI for the Choquet integral. *IEEE Transactions on Emerging Topics in Computational Intelligence*, pages 1 – 10, 2020. doi:10.1109/TETCI.2020.3005682.

[24] M. Sugeno. Theory of fuzzy integrals and applications, phd thesis. *Tokyo Inst. of Technology*, 1974.

[25] M. Sugeno, K. Fujimoto, and T. Murofushi. A hierarchical decomposition of Choquet integral model. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 3(1):1–15, 1995.

[26] J.-Z. Wu and G. Beliakov. Nonmodularity index for capacity identifying with multiple decision criteria. *Information Sciences*, 492:164–180, 2019.

Please cite this paper as: G. Beliakov, M. Gagolewski, S. James,
Hierarchical data fusion processes involving the Möbius representation of capacities,
Fuzzy Sets and Systems 433, 1–21, 2022, doi:10.1016/j.fss.2021.02.006