



**HAL**  
open science

# Learning Utilities and Sparse Representations of Capacities for Multicriteria Decision Making with the Bipolar Choquet Integral

Margot Herin, Patrice Perny, Nataliya Sokolovska

► **To cite this version:**

Margot Herin, Patrice Perny, Nataliya Sokolovska. Learning Utilities and Sparse Representations of Capacities for Multicriteria Decision Making with the Bipolar Choquet Integral. Multidisciplinary Workshop on Advances in Preference Handling, IJCAI, Jul 2022, Vienne, Austria. hal-03780464

**HAL Id: hal-03780464**

**<https://hal.science/hal-03780464v1>**

Submitted on 19 Sep 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Learning Utilities and Sparse Representations of Capacities for Multicriteria Decision Making with the Bipolar Choquet Integral

Margot Herin<sup>1</sup>, Patrice Perny<sup>1</sup>, Nataliya Sokolovska<sup>2</sup>

<sup>1</sup>LIP6 - Sorbonne University

<sup>2</sup>NutriOmics - Sorbonne University

{margot.herin,patrice.perny}@lip6.fr, nataliya.sokolovska@sorbonne-universite.fr

## Abstract

This paper concerns preference elicitation and learning of decision models in the context of multicriteria decision making. We propose an approach to learn a compact representation of preferences using a decomposable bi-polar Choquet integral. Our aim is to show how we can successively learn marginal utilities and then sparse representations of capacities involved in the decision model from preference data. We first present a preference elicitation method to learn spline representations of marginal utilities on every point of view. Then we propose a sparse learning approach for determining a compact representation of capacities fitted to the observed preferences. Finally, we present numerical tests showing the advantages of our approach compared to basic methods that do not seek sparsity or that force sparsity a priori by requiring k-additivity. We also apply our approach in the context of preferences over sets.

## 1 Introduction

Evaluation and decision making is often a matter of finding the most appropriate tradeoff between multiple and possibly conflicting points of views [Roy, 1996]. In the field of multicriteria decision making, various evaluation and aggregation models have been proposed to evaluate and compare the alternatives of a decision problem [Grabisch *et al.*, 2009]. These models generally use and combine objective and subjective information: on the one hand alternatives are described by consequence vectors representing their outcomes with respect to multiple points of views under consideration in the analysis of preferences. On the other hand, in order to go beyond straightforward preferences induced by Pareto dominance, more subjective preference parameters are used to model the value system of the Decision Maker (DM), e.g., the relative importance of the various points of views and their possible interactions in the evaluation process. Thus, a body of increasingly complex decision models is studied in decision theory to encompass an ever more sophisticated set of decision behaviors. This effort motivated by descriptive objectives comes at the cost of an additional complexity, both at the

level of preference learning (fitting the parameters of the preference model to the DM value system to explain or predict her preferences) and at the recommendation level (finding an optimal alternative becomes computationally more difficult). In this paper, we address the first challenge and propose a methodology dedicated to the identification of utilities and capacities in decision models involving Choquet integrals.

The Choquet integral is a well known aggregation function used in multicriteria decision making to assign an overall score to any evaluation vector attached to an alternative [Grabisch, 1996]. It performs a kind of sophisticated weighted average where weights are defined for every subset of components. The Choquet integral is also used in machine learning to replace the linear function of variables which is commonly used in standard regression methods [Gagolewski *et al.*, 2019; Beliakov and Divakov, 2020]. For example, logistic regression was extended to Choquistic regression [Tehrani *et al.*, 2011]. It is also used for learning to rank with the Choquet integral [Tehrani *et al.*, 2012] where the data are provided with the labels which are preference degrees from an ordered categorical scale.

The CIU model (Choquet Integral of Utilities) is a decision model where the Choquet integral is used to aggregate the utilities of alternatives with respect to  $n$  points of views. It is based on two types of preference parameters: utility functions defining the attractiveness of consequences on every relevant point of view and a set function named *capacity*, monotonic with respect to set inclusion, assigning a weight to every subset of points of views. The Choquet integral was initially introduced in the framework of decision under uncertainty [Schmeidler, 1989]; it has been generalized to be applied in multicriteria analysis [Grabisch, 1996; Grabisch and Labreuche, 2010] and some methods based on regression have been proposed for capacity identification, assuming the utilities are known [Grabisch *et al.*, 2008].

Here we consider also the bi-Choquet integral of utilities (bi-CUI) which is an extension of CIU using two capacities that cooperate in weighting criteria or subset of criteria; one applies to the positive part of the evaluation vector whereas the other applies to the negative part [Labreuche and Grabisch, 2006]. This extension inspired by Kahneman and Tversky's cumulative prospect theory (CPT) [Tversky and Kahneman, 1979] allows the representation of decision behaviors that may vary depending on whether positive or negative con-

sequences come into play. CPT was initially introduced in the context of decision making under risk and assumes that capacities are defined as monotone transformed of probability measures. The bi-CIU model under consideration here is more general than CPT by allowing any kind of (monotonic) capacity to weight the subsets of points of views [Labreuche and Grabisch, 2006]. It can be used in multicriteria optimization when criteria scales and preferences are bipolar [Martin and Perny, 2020]. The bi-CIU model can be further generalized using bi-capacities in Choquet integrals [Labreuche and Grabisch, 2006] but this latter generalization is not considered here for the sake of simplicity.

Our focus on CIU and bi-CIU models is motivated by several reasons: first CIU is acknowledged as one of the most general monotone compromise aggregators since it includes various simpler decision models as special cases (e.g., additive utilities, weighted sums, OWA and WOA aggregators [Yager, 1988; Torra, 1997]). Therefore CIU includes a rich family of aggregation functions which provides a natural setting to study how model complexity can be tuned to the preference system we want to describe or implement. Moreover the use of possibly non-additive capacities in CIU may require the definition of  $2^n$  parameters in the worst case (one for every subsets of points of view) where  $n$  is the number of points of view under consideration. This perfectly illustrates how the gain of expressiveness comes with an increase of model complexity and obviously raises the question of the parsimonious learning of the parameters defining the capacity. Then, considering bi-CIU is even more general and more powerful than CIU from a descriptive viewpoint. The bipolar version of the model being based on two capacities, it requires  $2^{n+1}$  weighting parameters, beside utility functions, which raises even more crucially the need of methods to learn sparse representations of capacities. It might indeed prevent over-fitting of preference data and lead to more compact and more explainable decision models.

For this reason, we study the potential of sparse learning to determine compact representations of preferences from preference data with the CIU or bi-CIU model. This problem is challenging due to the interplay of utilities and capacities in the computation of CIU values, making the learning of these two types of parameters interdependent. Another challenge comes from the fact that utilities and capacities are not directly observable and must be derived from preference statements over pairs of alternatives. We will propose an approach to learn utility functions and capacities in two successive steps: by properly selecting a first set of preference queries we will learn a spline representation of the utility function on every point of view and then we learn a sparse representation of the capacities from a database of preference examples.

The paper is organized as follows: Section 2 introduces the CIU and bi-CIU models and some related concepts. In Section 3 we present an elicitation approach to learn marginal utility functions defined on each point of view. In Section 4 we propose an approach to learn sparse capacities both for positive and negative consequences. In Section 5, we present numerical tests to compare the performances of our elicitation approach compared to baseline methods. Finally, we present in Section 6 an application to preference learning over sets.

## 2 Background on CIU and bi-CIU

We adopt the standard setting and notations for multiattribute or multicriteria decision making. Let  $N = \{1, \dots, n\}$  be the set of the  $n$  points of view to be considered in a decision problem. Let  $X = X_1 \times \dots \times X_n$  be the  $n$ -dimensional evaluation space where  $X_i$  is a bounded set of consequences ordered by relation  $\succ_i$ . For any  $i \in N$ , for any pair  $(x_i, y_i) \in X_i$ ,  $x_i \succ_i y_i$  (resp.  $x_i \succeq_i y_i$ ) means that  $x_i$  is a better consequence than  $y_i$  (resp. better or equal). Within every set  $X_i$  we distinguish three reference elements denoted  $-1_i$ ,  $0_i$  and  $1_i$  representing the bottom level, the neutral level and the top level consequences respectively [Grabisch and Labreuche, 2010]. These levels must be obtained in close cooperation with the DM. The alternatives to be compared are seen as elements of  $X$ . Thus, every alternative  $x \in X$  is described by its consequence vector  $(x_1, \dots, x_n)$  where  $x_i \in X_i$  is the consequence of  $x$  w.r.t. viewpoint  $i$ , for  $i = 1, \dots, n$ . In this setting we consider  $n$  utility functions  $u_i$  defined on  $X_i$  and strictly increasing with preference  $\succ_i$  for  $i = 1, \dots, n$ , such that  $u_i(-1_i) = -1$ ,  $u_i(0_i) = 0$  and  $u_i(1_i) = 1$ . Utilities are used to quantify the attractiveness of consequences on a common scale  $[-1, 1]$ . Consequences above the neutral level receive a positive utility whereas consequences below the neutral level receive a negative utility.

### 2.1 CIU and bi-CIU models

We recall here the definition of models CIU and bi-CIU that use a Choquet integral to aggregate the utilities defined above. Let  $v$  denote a capacity defined on  $2^N$ , i.e., a set function such that  $v(\emptyset) = 0$ ,  $v(N) = 1$  and  $v(A) \leq v(B)$  for all  $A, B \subseteq N$  such that  $A \subseteq B$ . The CIU model combines utilities  $u_i$ ,  $i = 1, \dots, n$  and the capacity  $v$  to define the value of any consequence vector  $x = (x_1, \dots, x_n)$  by the discrete Choquet integral of the utility vector  $u(x) = (u_1(x_1), \dots, u_n(x_n))$ . Formally, the CIU model reads as follows:

$$C_v(u(x)) = \sum_{i=1}^n [v(X_{(i)}) - v(X_{(i+1)})] u_{(i)}(x_{(i)}) \quad (1)$$

$$= \sum_{i=1}^n [u_{(i)}(x_{(i)}) - u_{(i-1)}(x_{(i-1)})] v(X_{(i)}) \quad (2)$$

where  $(\cdot)$  is any permutation of  $N$  such that  $u_{(i)}(x_{(i)}) \leq u_{(i+1)}(x_{(i+1)})$  and  $X_{(i)} = \{j \in N : u_{(j)}(x_{(j)}) \geq u_{(i)}(x_{(i)})\}$ ,  $i \in N$  with  $u_{(0)}(x_{(0)}) = 0$  and  $X_{(n+1)} = \emptyset$ .

**Example 1.** If  $N = \{1, 2, 3\}$  and  $u_2(x_2) \leq u_1(x_1) \leq u_3(x_3)$  then  $C_v(u(x_1, x_2, x_3)) = u_2(x_2)v(\{1, 2, 3\}) + [u_1(x_1) - u_2(x_2)]v(\{1, 3\}) + [u_3(x_3) - u_1(x_1)]v(\{3\})$  by Equation 2.

Then the preferences induced by CIU are obviously defined as follows: for any solutions  $x, y \in X$ ,  $x$  is at least as good as  $y$  (denoted  $x \succeq y$ ) if and only if  $C_v(u(x)) \geq C_v(u(y))$ . Similarly,  $x$  is indifferent to  $y$  (denoted  $x \sim y$ ) if and only if  $C_v(u(x)) = C_v(u(y))$ . Let us recall that monotonicity of  $v$  w.r.t. set inclusion and the monotonicity of  $u_i$  w.r.t.  $\succ_i$  are assumed to make sure that  $C_v(u(x)) \geq C_v(u(y))$  when  $x_i \succeq_i y_i$  for all  $i \in N$  (monotonicity of preference w.r.t. Pareto dominance).

Now, we consider the bi-CIU model that relies on the same utility functions than CIU but uses two capacities:

$$f_{v,w}^u(x) = C_v(u(x)^+) + C_w(-u(x)^-) \quad (3)$$

where  $u(x)^+$  (resp.  $u(x)^-$ ) is the utility vector  $u(x)$  (resp.  $-u(x)$ ) in which negative components are replaced by 0. It is well known that  $C_w(z) = -C_{\bar{w}}(-z)$  for any utility vector  $z$  where  $\bar{w}$  is the dual capacity of  $w$  defined by  $\bar{w}(A) = 1 - w(N \setminus A)$  for all  $A \subseteq N$ . Therefore  $f_{v,w}^u(x) = C_v(u(x)^+) - C_{\bar{w}}(u(x)^-)$ . This latter formulation makes more explicit the balance between positive and negative arguments like in cumulative prospect theory. Moreover, if  $v = w$ ,  $f_{v,w}^u(x) = C_v(u(x)^+) + C_v(-u(x)^-) = C_v(u(x))$  and therefore the bi-CIU model boils down to CIU.

**Example 1 (continued)** If  $u_2(x_2) \leq 0 \leq u_1(x_1) \leq u_3(x_3)$  then we have  $f_{v,w}^u(x_1, x_2, x_3) = C_v(u_1(x_1), 0, u_3(x_3)) + C_w(0, u_2(x_2), 0)$ . Equivalently  $f_{v,w}^u(x_1, x_2, x_3)$  also reads as follows:  $C_v(u_1(x_1), 0, u_3(x_3)) - C_{\bar{w}}(0, -u_2(x_2), 0)$ .

## 2.2 Möbius inverse and sparsity

An alternative representation of capacities and the Choquet integral relies on the Möbius inverse of the capacity. The Möbius inverse of  $v$  is another set function  $m_v$  defined on  $N$  by:  $m_v(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} v(B)$  for all  $A \subseteq N$ . The coefficients  $m_v(A)$  are called Möbius masses, they completely characterize  $v$ . We indeed have  $v(A) = \sum_{B \subseteq A} m_v(B)$ . The values of  $m_v$  can be positive or negative but add up to 1 since  $\sum_{B \subseteq N} m_v(B) = v(N) = 1$ .

Interestingly enough, the CIU model can be directly expressed from the Möbius inverse by:

$$C_v(u(x)) = \sum_{B \subseteq N} m(B) \min_{i \in B} \{u_i(x_i)\} \quad (4)$$

This formulation suggests that  $C_v(u(x))$  might admit a compact representation when the Möbius inverse is sparse (i.e., when the Möbius masses include many zeros or values that will not significantly impact the calculation). A frequent option used to obtain capacities having a sparse representation is to require that Möbius masses vanish for all subsets of states larger than a given  $k$  smaller than  $n$ . In this case, the resulting capacity is said to be  $k$ -additive [Grabisch, 1997]. For instance, when the capacity is 1-additive then all Möbius masses are null except for singletons where they are positive due to monotonicity. In this case, Equation 4 shows that *CIU* boils down to a simple additive utility function.

Considering only 2-additive capacities is a standard option to allow pairwise interactions while keeping a sparse model. For example, some learning methods have been proposed to simultaneously learn utilities and 2-additive capacities [Tehrani *et al.*, 2014; Galand and Mayag, 2017; Bresson *et al.*, 2020]. One may also wish to relax 2-additivity for  $k$ -additivity ( $k > 2$ ) with the aim of finding a better trade-off between sparsity and expressivity. However reasoning about sparsity in terms of  $k$ -additivity is a drastic reduction that may significantly impact our ability to fit preference data with relevant CIU models. It may indeed happen that very sparse but still  $n$ -additive capacities perfectly match preference data as shown hereafter:

**Example 2.** Let us consider a DM adopting an egalitarian attitude in the aggregation (focusing on the worse consequence) refined by an utilitarian criterion (using the sum of utilities to break ties). Such a decision attitude can be obviously represented by the  $\epsilon$ -min model  $f_\epsilon(x) = (1 - \epsilon) \min_{i \in N} u_i(x_i) + \frac{\epsilon}{n} \sum_{i=1}^n u_i(x)$  where  $\epsilon$  is a positive quantity chosen arbitrarily small. Clearly function  $f$  is an instance of CIU (see Equation 4) obtained for a capacity  $v$  whose Möbius inverse  $m_v$  is everywhere 0 excepted on singletons and on  $N$  ( $m_v(\{i\}) = \frac{\epsilon}{n}$  for all  $i \in N$  and  $m_v(N) = 1 - \epsilon$ ). We remark that due to the monotonicity w.r.t set inclusion, function  $v$  is non-null on every subset since Möbius masses are positive and non-null on singletons. Despite the fact that  $v$  is never null, it admits a very sparse representation in terms of Möbius masses where only  $n + 1$  out of  $2^n$  coefficients are non-null.

In the above Example, we remark that the most important Möbius mass is put on  $N$ , which shows that preferences induced by  $f_\epsilon$  could not be properly described by any  $k$ -additive capacity with  $k < n$  despite the fact that it can be closely approximated by the min model involving a single non-null Möbius mass (attached to  $N$ ). There is a need of new approaches to find sparse representations of capacities that best fit observed preferences, regardless  $k$ -additivity. In this paper we propose a more general approach to seek sparse Möbius representations of capacities and use it to learn simple instances of the CIU or bi-CUI model that best fit the preference data. We first present the learning of utility functions  $u_i$  in Section 3. The learning of sparse representations of capacities will be presented in Section 4.

## 3 Utility Elicitation

In order to elicit utility functions we use indifference statements between carefully selected alternatives to obtain useful constraints restricting the set of admissible utilities independently of the capacity. More precisely our approach consists in adapting the tradeoff method [Wakker and Deneffe, 1996; Abdellaoui, 2000] initially introduced in the context of cumulative prospect theory to the case of multicriteria evaluation with bipolar scales to elicit utility functions  $u_i, i = 1, \dots, n$ . Let  $i$  be any element of  $N$ . We present below the elicitation process to derive constraints on  $u_i$  below the neutral level  $\mathbf{0}_i$  and then above the neutral level  $\mathbf{0}_i$ . The preference queries involve tradeoffs between  $i$  and another element  $j$  of  $N$  that can be freely chosen.

### 3.1 Utility elicitation below the neutral level

Let  $r_j, R_j \in X_j$  and  $x_i \in X_i$  such that  $\mathbf{0}_j \prec_j r_j \prec_j R_j$ , i.e.,  $0 < u_j(r_j) < u_j(R_j)$  and  $x_i \prec_j \mathbf{0}_i$  i.e.,  $u_i(x_i) < 0$ . We consider the following query:

$Q_j^-(x_i)$  : what is the consequence  $y_i$  such that  $(x_i, r_j, \mathbf{0}_{-ij}) \sim (y_i, R_j, \mathbf{0}_{-ij})$ ?

where  $(x_i, r_j, \mathbf{0}_{-ij})$  is a vector of neutral consequences everywhere excepted on components  $i$  and  $j$  where values are  $r_j$  and  $x_i$ . The existence of an answer  $y_i \in X_i$  achieving the exact indifference requires a certain richness in  $X_i$  (solvability assumption). A case where the solvability assumption is not met is considered in Section 6.

We couple the observed indifference with a second one associated to the answer  $z_i$  to  $Q_j^-(h_i)$  for some  $h_i$  element of  $X_i \setminus \{h_i\}$ . Assuming  $(-\mathbf{1}_i, \mathbf{0}_{-i}) \prec \mathbf{0}$ , i.e.,  $w(N \setminus \{i\}) < 1$ , these indifferences imply the following equation as shown in Subsection 8.1 in the appendix:  $u_i(x_i) - u_i(y_i) = u_i(h_i) - u_i(z_i)$ . By taking  $h_i = y_i$ , we can obtain the following simplified equation:

$$u_i(x_i) - u_i(y_i) = u_i(y_i) - u_i(z_i) \quad (5)$$

### 3.2 Utility elicitation above the neutral level

Let  $r_j, R_j \in X_j$  and  $x_i \in X_i$  such that  $r_j \prec_j R_j \prec_j \mathbf{0}_j$ , i.e.,  $u_j(r_j) < u_j(R_j) < 0$  and  $x_i \succ_i \mathbf{0}_i$ , i.e.,  $u_i(x_i) > 0$ . We consider the following query:

$Q_j^+(x_i)$  : what is the consequence  $y_i$  such that  $(x_i, R_j, \mathbf{0}_{-ij}) \sim (y_i, r_j, \mathbf{0}_{-ij})$ ?

Similarly to the elicitation under the neutral level, we couple this observed indifference with a second one associated to the answer  $z_i$  to a question  $Q_j^+(h_i)$  for any  $h_i$  element of  $X_i \setminus \{x_i\}$ . Assuming  $(\mathbf{1}_i, \mathbf{0}_{-i}) \succ \mathbf{0}$ , i.e.,  $v(\{i\}) > 0$ , these indifferences imply the following equation as shown in Subsection 8.2 in the appendix:  $u_i(y_i) - u_i(x_i) = u_i(z_i) - u_i(h_i)$ . By taking  $h_i = y_i$ , we can obtain the following simplified equation:

$$u_i(y_i) - u_i(x_i) = u_i(z_i) - u_i(y_i) \quad (6)$$

### 3.3 Construction of utility curves

In the original method  $Q^+$  queries are involved in a recursive procedure known as *standard sequence* aiming at constructing a sequence of points on the utility curve  $u_i$ . The sequence  $(s_0, \dots, s_q)$  is constructed as follows:  $s_0 = \mathbf{0}_i$  and  $s_{t+1}$  is the answer to the query  $Q_j^+(s_t)$ . By construction this sequence is such that  $s_t \prec_i s_{t+1}$  and this improving sequence stops at step  $q$  when consequence  $\mathbf{1}_i$  is reached. Then we have  $u_i(s_{t+1}) - u_i(s_t) = u_i(s_t) - u_i(s_{t-1})$  by Equation 6, yielding the recursive equation:

$$u_i(s_{t+1}) = 2u_i(s_t) - u_i(s_{t-1}) \quad (7)$$

which completely determines the sequence, since  $u_i(s_0) = 0$  and  $u_i(s_q) = 1$ . We indeed have  $u_i(s_t) = t/q$  for  $t = 1, \dots, q$ . A symmetric sequence can be implemented to construct points on the utility curve below the neutral level using  $Q^-$  queries.

Such a method is known to be extremely sensitive to errors in the responses [Blavatsky, 2006]. Indeed, if one considers that every answer is provided with some noise such that  $u_i(s_{t+1}) = 2u_i(s_t) - u_i(s_{t-1}) + \epsilon_t$  for  $t = 1, \dots, q$  where  $\epsilon_t \sim \mathcal{U}([- \epsilon_{max}, \epsilon_{max}])$  and  $\epsilon_{max} > 0$ , the estimation error on the points of the standard sequence can be large. Indeed  $u_i(s_t) = \frac{t}{q} + \sum_{k=1}^t \epsilon_k - \frac{t}{q} \sum_{k=1}^q \epsilon_k$  for  $t = 1, \dots, q$  and the expected squared estimation error on a point  $s_t$  is:  $\mathbb{E}[(u_i(s_t) - t/q)^2] = \mathbb{E}[(\sum_{k=1}^t \epsilon_k)^2] + t^2/q^2 \mathbb{E}[(\sum_{k=1}^q \epsilon_k)^2] - \frac{2t}{q} \mathbb{E}[\sum_{k_1=1}^t \sum_{k_2=1}^q \epsilon_{k_1} \epsilon_{k_2}] = \frac{2\epsilon_{max}^2}{12} (t - \frac{t^2}{q})$ .

Then the maximum expected estimation error is reached on the middle of the sequence and is proportional to the length  $q$  of the sequence and to the squared noise intensity  $\epsilon_{max}^2$ :

$$\max_t \mathbb{E}[(u_j(s_t) - t/q)^2] = q\epsilon_{max}^2/36$$

Since the error increases with the length of the standard sequence, we propose to consider multiple minimal length ( $q = 2$ ) standard sequences of type  $(s_0, s_1, s_2)$ . Multiplicity is obtained by varying the initial location  $s_0$ , the reference dimension  $j$  and the mesh  $(r_j, R_j)$ . Note that if  $s_0$  is below the neutral level we use decreasing sequences generated by  $Q^-$  queries. The resulting triplets  $(s_0^k, s_1^k, s_2^k)$  yield a set of linear constraints on  $u_i$  given by Equation 6. Then we define  $u_i$  as a I-spline function that best fits the resulting set of equalities.

Spline functions are piecewise polynomial functions of class  $C^k$  widely used for data interpolation or approximation due to their ability to smoothly approximate complex shapes. Moreover they allow for a compact representation of utilities. Indeed, a spline function can be expressed as a linear combination of basis functions and is thus characterized by the coefficients of the combination. Since utility functions are supposed increasing, we will use a basis  $(I_l)_{l=1}^L$  of monotonically increasing spline functions, known as I-spline functions [Ramsay, 1988] weighted by positive coefficients (adding up to 1 so as to have  $u_i(\mathbf{1}_i) = 1$ ). We use here cubic I-splines ( $k = 3$ ) because they have matching first and second derivatives while preserving a local influence of every components. Note that we use a translation of the basis functions from  $[0; 1]$  to  $[-1; 1]$ . Formally,  $u_i$  is defined by parameters  $\alpha_i = (\alpha_{i,1}, \dots, \alpha_{i,L}) \in [0, 1]^L$  such that:

$$\forall x_i \in [-\mathbf{1}_i, \mathbf{1}_i], \quad u_i(x_i) = 2 \sum_{l=1}^L \alpha_{i,l} I_l(x) - 1 \quad (8)$$

Using Equation 8, the problem of finding the utility that best fits the equalities can be formalized as a linear program with relaxed constraints:

$$\begin{aligned} & \min \sum_k (\epsilon_k^+ + \epsilon_k^-) \\ & \begin{cases} \sum_{l=1}^L \alpha_{i,l} (2I_l(s_0^k) - I_l(s_1^k) - I_l(s_2^k)) + \epsilon_k^+ - \epsilon_k^- = 0 \\ 2 \sum_{l=1}^L I_l(\mathbf{1}_i) \alpha_{l,j} - 1 = 1 \\ 2 \sum_{l=1}^L I_l(-\mathbf{1}_i) \alpha_{l,j} - 1 = -1 \\ 2 \sum_{l=1}^L I_l(\mathbf{0}_i) \alpha_{l,j} - 1 = 0 \end{cases} \\ & \epsilon_k^+ \geq 0, \epsilon_k^- \geq 0, \alpha_{i,l} \geq 0. \end{aligned}$$

## 4 Sparse learning of capacities

We present the learning of sparse representations of capacities directly on the bi-CIU model because it includes the CIU model as special case. Let us reformulate bi-CIU from the Möbius inverse of capacities  $v$  and  $w$ . We obtain:  $f_{v,w}^u(x) = \sum_{B \subseteq N} m_v(B) \min_{i \in N} \{u_i(x_i)^+\} + \sum_{B \subseteq N} m_w(B) \min_{i \in B} \{-u_i(x_i)^-\}$ .

Given a set of preference statements  $\mathcal{P} = \{(x^p, y^p) \in \mathcal{X}^2 : x^p \succeq y^p, \text{ or } x^p \sim y^p = 1, \dots, N\}$ , and a utility function  $u$ , one wants to find the capacities  $v, w$  that match with the observed preferences. We also want to obtain a sparse Möbius representation of  $v$  and  $w$  through  $L_1$  penalization. However, because of the constraints  $v(N) = \sum_B m_v(B) = 1$  and  $w(N) = \sum_B m_w(B) = 1$ , the effect of the penalization term on the  $L_1$  norm of the Möbius masses is cancelled at some point, since  $\|m_v\|_1 \geq \sum_{B \subseteq N} m(B) = 1$  and  $\|m_w\|_1 \geq \sum_{B \subseteq N} m(B) = 1$ . As a solution, we use the following representation of  $m_v$  and  $m_w$  :  $m_v(B) = 1/n + c_v(B)$ ,

$m_w(B) = 1/n + c_w(B)$  when  $|B| = 1$  and  $m_v(B) = c_v(B)$ ,  $m_w(B) = c_w(B)$  otherwise. Then we penalize on  $\|c_v\|_1$  and  $\|c_w\|_1$  instead of  $\|m_v\|_1$  and  $\|m_w\|_1$ . This leads to solve the following sparse constraint-based regression (written in its linearized form):

$$\begin{aligned} & \min \sum_p (\epsilon_p^+ + \epsilon_p^-) + \sum_p \epsilon_p + \lambda \sum_{B \subseteq N} (a_v(B) + b_v(B)) \\ & + \lambda \sum_{B \subseteq N} (a_w(B) + b_w(B)) \\ & \begin{cases} \text{if } x^p \sim y^p, \\ \sum_{B \subseteq N} m_v(B) U_p^+(B) + m_w(B) U_p^-(B) + \epsilon_p^+ - \epsilon_p^- = 0 \\ \text{if } x^p \succeq y^p, \\ \sum_{B \subseteq N} m_v(B) U_p^+(B) + m_w(B) U_p^-(B) + \epsilon_p \geq 0 \\ m_v(B) = 1/n + c_v(B), \quad \forall B \subseteq N : |B| = 1 \\ m_v(B) = c_v(B), \quad \forall B \subseteq N : |B| > 1 \\ m_w(B) = 1/n + c_w(B), \quad \forall B \subseteq N : |B| = 1 \\ m_w(B) = c_w(B), \quad \forall B \subseteq N : |B| > 1 \\ c_v(B) = a_v(B) - b_v(B), \quad \forall B \subseteq N \\ c_w(B) = a_w(B) - b_w(B), \quad \forall B \subseteq N \\ \sum_{A \subseteq B} m_v(A \cup \{i\}) \geq 0, \quad \forall i \in N, \quad \forall B \subseteq N \setminus \{i\} \\ \sum_{A \subseteq B} m_w(A \cup \{i\}) \geq 0, \quad \forall i \in N, \quad \forall B \subseteq N \setminus \{i\} \end{cases} \\ & \epsilon_p^+ \geq 0, \epsilon_p^- \geq 0, \epsilon_p \geq 0, a_v(B) \geq 0, b_v(B) \geq 0, a_w(B) \geq 0, \\ & b_w(B) \geq 0, m_v(B), m_w(B), c_v(B), c_w(B) \in \mathbb{R} \end{aligned}$$

## 5 Experiments

In this section we show the results of numerical experiments on synthetic data and we illustrate the advantage of our approach over the baseline methods.

### 5.1 Synthetic data

We generate the synthetic data as follows. First, random bipolar sparse Choquet models  $f_{v,w}^u$  are created through the generation of a  $n$  utility function  $u_i$ . Second, we generate  $m_v$  and  $m_w$ , two sparse Möbius representations of capacities. For this, the degree of sparsity  $d_s$  (percentage of null Möbius masses) is fixed, and two capacities with  $d_s$ -sparse Möbius representations are generated randomly without requiring for monotonicity. Then  $m_v, m_w$  are taken as the Möbius representations of the closest (in the sense of the  $L_1$  norm) capacities verifying the monotonicity property.

Third, for  $n$  given utility functions  $u_i$  and capacities  $v, w$ , we simulate Q-queries and their answers for the utility learning and preferences compatible with  $f_{v,w}^u$  for the capacity learning. Answers to Q-queries are provided with some random uniform noise  $\epsilon \in [-\epsilon_{max}, \epsilon_{max}]$ . For the capacity learning, pairs  $(x^p, y^p)$  of objects taking values on  $n$  criteria are drawn uniformly within  $[-1_i, 1_i], i = 1 \dots n$ . In order to introduce noise, each example is associated with a preference statement  $x^p \preceq y^p$  or  $x^p \succeq y^p$  if  $|(f_v^u(x^p) + \sigma_x) - (f_v^u(y^p) + \sigma_y)| \geq \sigma$ , and  $x^p \sim y^p$  otherwise, where  $\sigma_x, \sigma_y$  are noise values uniformly taken within an interval  $[-\sigma, \sigma]$ . Pairs with Pareto dominance are discarded. This process is used to generate training sets of size  $N$  which we vary in our experiments, and test sets of size 10000. In the following, the generalizing performance of the models is evaluated as the percentage of preference error in a test set (test error).

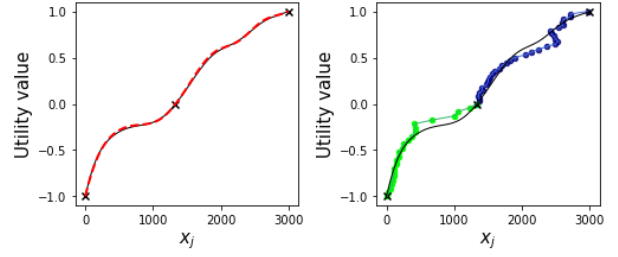


Figure 1: Learned utility function with our method (left: red dotted line) and with standard sequences (right: green and blue points) along with the ground truth (black plain line).

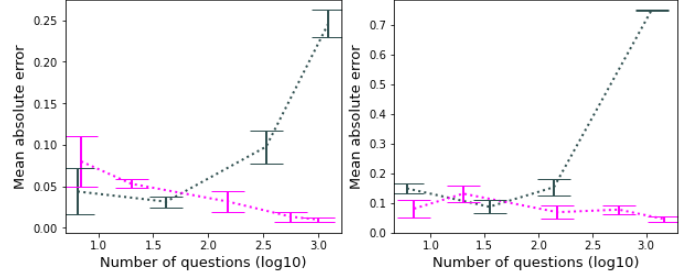


Figure 2: Average MAE w.r.t the number of questions for our method (fushia) and standard sequences (grey) over 10 simulations for  $\epsilon_{max} = 0.05$  (left) and  $\epsilon_{max} = 0.1$  (right)

### 5.2 Utility learning

We conduct numerical tests on utilities learning. First, we generate a sparse model  $f_{v,w}^u$  with sparsity degree  $d_s = 0.95$  and learn the utility function  $u_1$  with standard sequences and then with our method. Answers to Q-queries are simulated with a level of noise  $\epsilon_{max} = 0.05$ . Figure 1 displays the estimated utility functions for both methods along with the ground truth  $u_i$ . On the left, the estimation provided by our method perfectly matches the ground truth while on the right the estimation of the standard sequence clearly suffers from noise distortion. We conducted the same experiment on 10 random sparse models, and obtained an average mean absolute error (MAE) of  $0.084 \pm 0.052$  for the standard sequence method and  $0.022 \pm 0.008$ . The MAE is the mean absolute difference between the ground truth and the estimated values on a fixed subdivision.

On Figure 2 we represent the accuracy of both methods in terms of MAE as a function of the number of questions asked, we show the average values for 10 random sparse models. Figure 2 shows the case  $\epsilon_{max} = 0.05$  on the left and  $\epsilon_{max} = 0.1$  on the right. The test confirms that long standard sequences lead to very poor results. On the other hand, one can see that regardless the level of noise, our approach converges to a null MAE. This second method seems to be a more robust approach.

### 5.3 Learning Möbius representations

We first illustrate the process of learning a sparse Möbius capacity representation in the specific case of Example 2, and then we proceed to simulations to demonstrate the benefits of

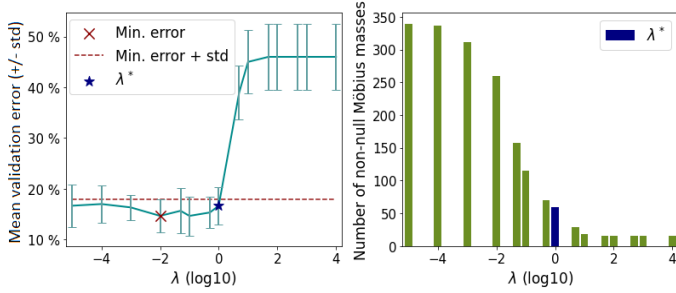


Figure 3: Selection of the hyperparameter  $\lambda$  with cross validation: mean percentage test error on the 5 tests folds (left) and  $L_0$ -norm of the learned models according  $\lambda$  (right).

our approach in the general case of arbitrarily sparse data.

### Illustration on Example 2

We generate preference data according to the model of Example 2 ( $\epsilon$ -min model) with  $n = 8$ ,  $N = 500$  and  $\sigma = 0.03$  (which corresponds to approximately 10% of preference error in the training set). We compare our method based on sparse regression to some baselines, such as the unpenalized regression and the use of unpenalized 2-additive models (2-ADD). The hyper-parameter  $\lambda^*$  for our approach is selected by cross validation with a number of folds equal to 5. The one-standard-error-rule is used, i.e., one selects  $\lambda^*$  as the highest value of  $\lambda$  that yields a mean error on the validation folds lower than the minimum mean error over all  $\lambda$  plus the standard error associated to this minimum. On the left of Figure 3 we show the average generalizing performance (mean validation error) of the learned models over the 5 folds of the training set for different  $\lambda$ ; the selected one-standard-error-rule  $\lambda^*$  is highlighted. One can observe (Figure 3 on the right) that the increase of  $\lambda$  increases the sparsity of the solution, and  $\lambda^*$  corresponds to the model with the highest compactness for a still competitive generalizing performance. On Figure 4 we illustrate the learned Möbius masses for  $\lambda^*$  (left), the unpenalized regression ( $\lambda = 0$ ) (middle) and 2-ADD (right) along with the ground truth ( $\epsilon$ -min model). It is clear that the regression without any penalty term fails to recover the  $\epsilon$ -min model; it does not find any compact representation either. It achieves, however, a reasonable generalizing performance on the test set (test error of 14.33%). The 2-additive model, while being compact, is far from the ground truth and does not capture high interactions, leading to a poor generalizing performance (test error of 39.67%). Our approach combines both advantages of the baselines: compactness and optimal generalizing performance (test error of 14.00%).

### Comparative performance on arbitrarily sparse models

We observed that the 2-additive models can fail if the underlying data structure contains richer interactions. In this section, we extend our comparisons to the use of  $k$ -additive models for  $k = 1, \dots, n - 1$ , and we search for an optimal  $k^*$  (chosen by cross-validation). We also compare the results of our method to the unpenalized regression method.

For two degrees of sparsity ( $d_s = 0.95$  and  $d_s = 0.90$ ) we generate 10 hidden arbitrarily sparse models with the same

$\ m\ _0$	Model	Test error	$\ \hat{m}\ _0$
$12.9 \pm 7.26$ ( $d_s = 0.95$ )	Sparse reg.	<b>12.1 <math>\pm</math> 2.5%</b>	78.6 $\pm$ 34.1
	Unp. reg.	12.8 $\pm$ 2.1%	329.4 $\pm$ 34.1
	4-ADD	13.5 $\pm$ 2.7%	235.9 $\pm$ 56.6
	2-ADD	20.7 $\pm$ 5.0%	<b>36.2 <math>\pm</math> 5.2</b>
	$k^*$ -ADD	14.4 $\pm$ 3.1%	159.2 $\pm$ 115.8
$30.5 \pm 11.08$ ( $d_s = 0.90$ )	Sparse reg.	<b>15.1 <math>\pm</math> 2.5%</b>	112.5 $\pm$ 99.8
	Unp. reg.	15.3 $\pm$ 2.3%	402.0 $\pm$ 48.3
	4-ADD	15.8 $\pm$ 0.7%	271.3 $\pm$ 33.3
	2-ADD	20.4 $\pm$ 2.3%	<b>41.2 <math>\pm</math> 5.2</b>
	$k^*$ -ADD	16.1 $\pm$ 2.6%	167.4 $\pm$ 113.9

Table 1: Generalizing performance and compactness of the proposed approach and the baselines depending on the level of sparsity of the hidden models. Results are averaged over 10 simulations.

parameters as in the previous experiment. The generalizing performance (test error) of our approach (sparse regression), the unpenalized regression (unp. reg.), 2-ADD, 4-ADD and  $k^*$ -ADD ( $k$ -additive where  $k$  has been chosen by cross-validation) are averaged and displayed in Table 1 along with the average sparsity of the learned models ( $\|\hat{m}\|_0$ ).

Our approach clearly outperforms the methods based on  $k$ -additive models in terms of generalizing performance with a competitive compactness of the obtained representations. Indeed, with the 2-ADD method, our method seems to be the only one that reaches a  $L_0$ -norm close to the one of the hidden models. Note that, while having a generalizing performance close to optimal, the unpenalized regression (unp. reg.) is unable to recover an underlying sparse model. On Figure 5 we show the evaluations for each method using both criteria: test error and the number of non-null Möbius masses. The diamond curve corresponds to the results of the sparse regression for different values of  $\lambda$  (and  $\lambda^*$ ), and the circle curve to  $k$ -additive method for different values of  $k$  (and  $k^*$ ). One can see that the sparse regression allows to achieve a better compromise between sparsity and accuracy compared to the method based on  $k$ -additive models.

## 6 Application: preferences over teams

In this section we consider the problem of learning preferences over teams seen as subsets of elements (individuals), each being assigned to a specific role (e.g., soccer team, project group). We consider a reference set  $E$  of elements. A team is characterized by a vector  $x = (x_1, \dots, x_n)$  where  $x_i \in E$  is the name of the element at position  $i$ . Preferences over teams are modeled by a bi-CIU model  $f_{v,w}^u$ . This model defines the value of a team as a function of the value of its elements (modelled by  $u$ ) and the importance of interactions between roles (modelled by  $v$  and  $w$ ). This is a particular case of the general setting of this paper where the  $X'_i$ 's are equal to  $E$ . The  $u'_i$ 's can be used to measure the values of the elements acting at positions  $i$  in the team but for the sake of simplicity they are assumed to be independent of  $i$  and denoted  $u$ . We assume that the relative value of elements is known and represented by a weak-order  $\succsim$  on  $E$  but that the utilities (monotonic w.r.t.  $\succsim$ ) are to be learned. Let  $\mathbf{0} \in E$

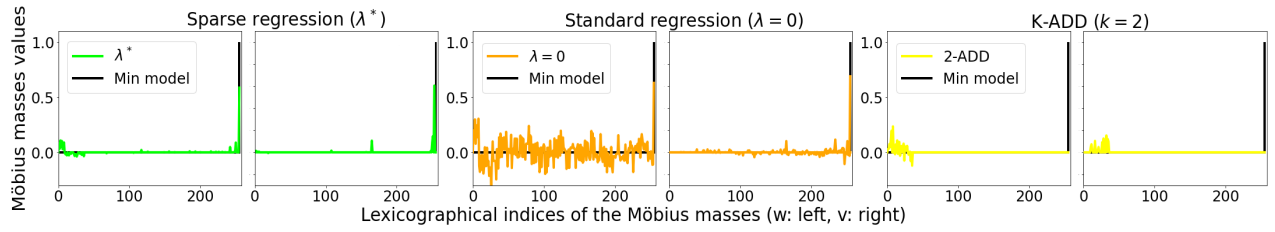


Figure 4: Learned models and hidden model (min model)

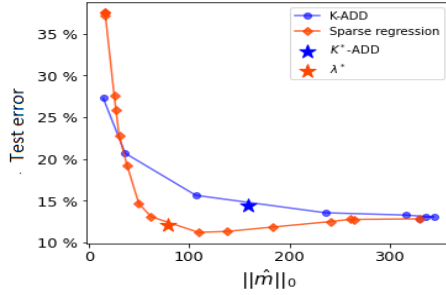


Figure 5: Compromise between generalizing performance and compactness achieved by the sparse regression and K-ADD.

(resp.  $-1$  and  $1$ ) denote any element of the indifference class of medium-level elements (resp. bottom-level and top-level elements). The elements of  $E$  are indexed in such a way that  $e_k \succsim e_{k+1}$ ,  $k = 1, \dots, |E| - 1$ . In this setting we apply our learning approach to derive a value model for teams from preference examples and to infer new preferences over new pairs of teams. As  $E$  is a finite set, the solvability assumption required the Q-queries might not hold and the elicitation process must be adapted. We propose the following procedure. For any  $i, j \in \{1 \dots n\}$ , inequalities constraints on  $u$  are obtained by asking two queries:

**Below the neutral level:** Let  $i, j \in \{1, \dots, n\}$ ,  $x_i, r_j, R_j \in E$  such that  $x_i \succsim \mathbf{0}$  and  $R_j \succ r_j \succ \mathbf{0}$ .

$LQ_j^-(x_i)$ : what is the element  $y_i^+$  of lowest index such that  $(x_i, r_j, \mathbf{0}_{-ij}) \succsim (y_i^+, R_j, \mathbf{0}_{-ij})$ ?

$HQ_j^-(y_i^+)$ : what is the element  $h_i^-$  of highest index such that  $(y_i^+, r_j, \mathbf{0}_{-ij}) \succsim (h_i^-, R_j, \mathbf{0}_{-ij})$ ?

Assuming  $(-\mathbf{1}_i, \mathbf{0}_{-i}) \prec \mathbf{0}$ , i.e.,  $w(N \setminus \{i\}) < 1$ , we obtain (see Subsection 8.3 in the appendix):

$$u(y_i^+) - u(h_i^-) \geq u(x_i) - u(y_i^+) \quad (9)$$

$$u(y_i^+) - u(h_i^+) > u(x_i) - u(y_i^-) \quad (10)$$

where  $y_i^+ = e_{k+1}$ ,  $y_i^- = e_k$  and  $h_i^+ = e_{k'+1}$ ,  $h_i^- = e_{k'}$  for some  $k, k' \in \{1, \dots, |E| - 1\}$ .

**Above the neutral level:** Let  $x_i, r_j, R_j \in E$  such that  $x_i \succ \mathbf{0}$  and  $r_j \prec R_j \prec \mathbf{0}$ :

$HQ_j^+(x_i)$ : what is the element  $y_i^-$  of highest index such that  $(x_i, R_j, \mathbf{0}_{-ij}) \succsim (y_i^-, r_j, \mathbf{0}_{-ij})$ ?

$LQ_j^+(y_i^-)$ : what is the element  $h_i^+$  of lowest index such that

$$(y_i^-, R_j, \mathbf{0}_{-ij}) \succsim (h_i^+, r_j, \mathbf{0}_{-ij})?$$

Assuming  $(\mathbf{1}_i, \mathbf{0}_{-i}) \succ \mathbf{0}$ , i.e.,  $v(\{i\}) > 0$ , we obtain (see Subsection 8.3 in the appendix):

$$u(y_i^-) - u(h_i^+) \leq u(x_i) - u(y_i^-) \quad (11)$$

$$u(y_i^-) - u(h_i^-) > u(x_i) - u(y_i^+) \quad (12)$$

Then the elicitation of  $u$  presented in Section 3 is adapted by replacing Equations 5 and 6 by Equations 9,10 and 11, 12. We obtain the following numerical result: for 10 random sparse models  $f_{v,w}^u$  ( $d_s = 0.95$ ,  $\epsilon_{max} = 0.05$ ,  $n = 7$ ,  $|E| = 200$ ), we obtain a mean absolute error on  $u$  of  $0.103 \pm 0.057$  (for  $115.8 \pm 18.84$  queries). This represents 5% of the range of the utility scale. The error is higher than the average error on  $u$  obtained in Subsection 5.2 for continuous attributes due to the fact that inequations are less informative than equations. The elicitation of  $v, w$  remains unchanged.

## 7 Conclusion

We proposed a new method to learn both utilities and sparse representations of capacities in the context of multicriteria decision making using a bipolar Choquet integral. It also applies to learn preference over sets. We first proposed a variant of the tradeoff method to learn utility functions in CUI and bi-CUI models which appears to be more robust than usual elicitation methods based on standard sequences. Then we presented a method to learn compact representations of capacities. The Lasso introduced by [Tibshirani, 1996] is a well studied approach to control the model's complexity. We are aware of several attempts to obtain sparse solutions for the non-additive integrals via the  $L_1$  penalty term. For example, the sparsity inducing penalty was applied to the capacity [Anderson *et al.*, 2014; Adeyeba *et al.*, 2015]; the penalised sum of squared errors with Gini-Simpson index regularisation and the  $L_0$  norm on the Shapley values were considered in [Pinar *et al.*, 2017]. The  $L_1$  penalty was also applied to capacities represented by interaction indices in [de Oliveira *et al.*, 2022]. We also explored the  $L_1$  penalty term to obtain a sparse solution, however, we focused on learning sparse Möbius representations, and we illustrate by our experiments that the proposed method outperforms the baselines such as an unregularised regression and  $k$ -additive models.

## Acknowledgments

We acknowledge a financial support from the project THEMIS ANR-20-CE23-0018 of the French National Research Agency (ANR).



## References

- [Abdellaoui, 2000] Mohammed Abdellaoui. Parameter-free elicitation of utility and probability weighting functions. *Management Science*, 46(11):1497–1512, 2000.
- [Adeyeba *et al.*, 2015] Titilope A. Adeyeba, Derek T. Anderson, and Timothy C. Havens. Insights and characterization of  $l_1$ -norm based sparsity learning of a lexicographically encoded capacity vector for the Choquet integral. In *FUZZ-IEEE*, pages 1 – 7, 2015.
- [Anderson *et al.*, 2014] Derek T. Anderson, Stanton R. Price, and Timothy C. Havens. Regularization-based learning of the Choquet integral. In *FUZZ-IEEE*, pages 2519 – 2526, 2014.
- [Beliakov and Divakov, 2020] Gleb Beliakov and Dmitriy Divakov. On representation of fuzzy measures for learning Choquet and Sugeno integrals. *Knowledge-Based Systems*, 189:105134, 2020.
- [Blavatsky, 2006] Pavlo Blavatsky. Error propagation in the elicitation of utility and probability weighting functions. *Theory and Decision*, 60(2):315–334, 2006.
- [Bresson *et al.*, 2020] Roman Bresson, Johanne Cohen, Eyke Hüllermeier, Christophe Labreuche, and Michèle Sebag. Neural representation and learning of hierarchical 2-additive Choquet integrals. In *IJCAI*, pages 1984–1991, 2020.
- [de Oliveira *et al.*, 2022] Henrique Evangelista de Oliveira, Leonardo Tomazeli Duarte, and João Marcos Travassos Romano. Identification of the Choquet integral parameters in the interaction index domain by means of sparse modeling. *Expert Systems with Applications*, 187, 2022.
- [Gagolewski *et al.*, 2019] Marek Gagolewski, Simon James, and Gleb Beliakov. Supervised learning to aggregate data with the Sugeno integral. *IEEE Transactions on Fuzzy Systems*, 27(4):810 – 815, 2019.
- [Galand and Mayag, 2017] Lucie Galand and Brice Mayag. A heuristic approach to test the compatibility of a preference information with a Choquet integral model. In *ADT*, pages 65 – 80, 2017.
- [Grabisch and Labreuche, 2010] Michel Grabisch and Christophe Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *Annals of Operations Research*, 175(1):247–286, 2010.
- [Grabisch *et al.*, 2008] Michel Grabisch, Ivan Kojadinovic, and Patrick Meyer. A review of methods for capacity identification in Choquet integral based multi-attribute utility theory: Applications of the Kappalab R package. *European journal of operational research*, 186(2):766–785, 2008.
- [Grabisch *et al.*, 2009] Michel Grabisch, Jean-Luc Marichal, Radko Mesiar, and Endre Pap. *Aggregation functions*, volume 127. Cambridge University Press, 2009.
- [Grabisch, 1996] Michel Grabisch. The application of fuzzy integrals in multicriteria decision making. *European journal of operational research*, 89(3):445–456, 1996.
- [Grabisch, 1997] Michel Grabisch. K-order additive discrete fuzzy measures and their representation. *Fuzzy sets and systems*, 92(2):167–189, 1997.
- [Labreuche and Grabisch, 2006] Christophe Labreuche and Michel Grabisch. Generalized choquet-like aggregation functions for handling bipolar scales. *European Journal of Operational Research*, 172(3):931–955, 2006.
- [Martin and Perny, 2020] Hugo Martin and Patrice Perny. New computational models for the choquet integral. In *ECAI 2020*, pages 147–154. IOS Press, 2020.
- [Pinar *et al.*, 2017] Anthony J. Pinar, Derek T. Anderson, Timothy C. Havens, Alina Zare, and Titilope Adeyeba. Measures of the Shapley index for learning lower complexity fuzzy integrals. *Granul. Comput.*, 2:303 – 319, 2017.
- [Ramsay, 1988] James O. Ramsay. Monotone regression spline in action. *Statistical Science*, page 425–441, 1988.
- [Roy, 1996] Bernard Roy. *Multicriteria methodology for decision aiding*, volume 12. Springer Science & Business Media, 1996.
- [Schmeidler, 1989] David Schmeidler. Subjective probability and expected utility without additivity. *Econometrica*, 57(3):571–587, 1989.
- [Tehrani *et al.*, 2011] Ali Fallah Tehrani, Weiwei Cheng, and Eyke Hüllermeier. Choquistic regression: generalizing logistic regression using the Choquet integral. In *EUSFLAT*, 2011.
- [Tehrani *et al.*, 2012] Ali Fallah Tehrani, Weiwei Cheng, and Eyke Hüllermeier. Preference learning using the Choquet integral: the case of multipartite ranking. *IEEE Transactions on Fuzzy Systems*, 20(6):1102 – 1113, 2012.
- [Tehrani *et al.*, 2014] Ali Fallah Tehrani, Christophe Labreuche, and Eyke Hüllermeier. Choquistic utilitarianistic regression. In *DA2PL*, pages 35 – 42, 2014.
- [Tibshirani, 1996] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (methodological)*, 58(1):267 – 88, 1996.
- [Torra, 1997] Vicenç Torra. The weighted owa operator. *International Journal of Intelligent Systems*, 12(2):153–166, 1997.
- [Tversky and Kahneman, 1979] Amos Tversky and Daniel Kahneman. An analysis of decision under risk. *Econometrica*, 47(2):263–292, 1979.
- [Wakker and Deneffe, 1996] Peter Wakker and Daniel Deneffe. Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management science*, 42(8):1131–1150, 1996.
- [Yager, 1988] Ronald R Yager. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on systems, Man, and Cybernetics*, 18(1):183–190, 1988.

## 8 Appendix

In this appendix we explain how equations on utilities (Equations 5–6 and Equations 9–12) are derived from preference and indifference statements.

### 8.1 Utility elicitation below the neutral level

Let  $r_j, R_j \in X_j$  and  $x_i \in X_i$  such that  $\mathbf{0}_j \prec_j r_j \prec_j R_j$ , i.e.,  $0 < u_j(r_j) < u_j(R_j)$  and  $x_i \prec_j \mathbf{0}_i$  i.e.,  $u_i(x_i) < 0$ . We consider the following query:

$Q_j^-(x_i)$  : what is the consequence  $y_i$  such that  $(x_i, r_j, \mathbf{0}_{-ij}) \sim (y_i, R_j, \mathbf{0}_{-ij})$ ?

From this indifference we have:  $f_{v,w}^u(x_i, r_j, \mathbf{0}_{-ij}) = f_{v,w}^u(y_i, R_j, \mathbf{0}_{-ij})$ . Since  $r_j \prec_j R_j$  we have  $y_i \prec_i x_i$ . Moreover, since  $x_i \prec_j \mathbf{0}_i$ , i.e.,  $u_i(x_i) \leq 0$ , we have  $u_i(y_i) < u_i(x_i) \leq 0 \leq u_j(r_j) < u_j(R_j)$  and therefore  $f_{v,w}^u(x_i, r_j, \mathbf{0}_{-ij}) = u_j(r_j)v(\{j\}) + u_i(x_i)(1 - w(N \setminus \{i\}))$ . Similarly  $f_{v,w}^u(y_i, R_j, \mathbf{0}_{-ij}) = u_j(R_j)v(\{j\}) + u_i(y_i)(1 - w(N \setminus \{i\}))$ . Hence we have:  $(u_i(x_i) - u_i(y_i))(1 - w(N \setminus \{i\})) = (u_j(R_j) - u_j(r_j))v(\{j\})$  (a).

Similarly, for any  $h_i$  element of  $X_i \setminus \{x_i\}$  chosen such that  $h_i \prec_i \mathbf{0}_i$ , i.e.,  $u_i(h_i) < 0$ , if  $z_i$  is the answer to question  $Q_j^-(h_i)$  we have:  $(u_i(h_i) - u_i(z_i))(1 - w(N \setminus \{i\})) = (u_j(R_j) - u_j(r_j))v(\{j\})$  (b).

From Equations (a) and (b) we have  $(u_i(x_i) - u_i(y_i))(1 - w(N \setminus \{i\})) = (u_i(h_i) - u_i(z_i))(1 - w(N \setminus \{i\}))$ . Assuming  $(-\mathbf{1}_i, \mathbf{0}_{-i}) \prec \mathbf{0}$ , i.e.,  $w(N \setminus \{i\}) < 1$  we obtain:

$$u_i(x_i) - u_i(y_i) = u_i(h_i) - u_i(z_i) \quad (13)$$

By taking  $h_i = y_i$ , we can obtain the following simplified equation:

$$u_i(x_i) - u_i(y_i) = u_i(y_i) - u_i(z_i)$$

### 8.2 Utility elicitation above the neutral level

Let  $r_j, R_j \in X_j$  and  $x_i \in X_i$  such that  $r_j \prec_j R_j \prec_j \mathbf{0}_j$ , i.e.,  $u_j(r_j) < u_j(R_j) < 0$  and  $x_i \succ_i \mathbf{0}_i$ , i.e.,  $u_i(x_i) > 0$ . We consider the following query:

$Q_j^+(x_i)$  : what is the consequence  $y_i$  such that  $(x_i, R_j, \mathbf{0}_{-ij}) \sim (y_i, r_j, \mathbf{0}_{-ij})$ ?

From this indifference we have:  $f_{v,w}^u(x_i, R_j, \mathbf{0}_{-ij}) = f_{v,w}^u(y_i, r_j, \mathbf{0}_{-ij})$ . Since  $r_j \prec_j R_j$  we have  $x_i \prec_i y_i$ . Moreover, since  $x_i \succ_i \mathbf{0}_i$ , i.e.,  $u_i(x_i) \geq 0$ , we have  $u_j(r_j) < u_j(R_j) \leq 0 \leq u_i(x_i) < u_i(y_i)$  and therefore  $f_{v,w}^u(x_i, R_j, \mathbf{0}_{-ij}) = u_i(x_i)v(\{i\}) + u_j(R_j)(1 - w(N \setminus \{j\}))$ . Similarly  $f_{v,w}^u(y_i, r_j, \mathbf{0}_{-ij}) = u_i(y_i)v(\{i\}) + u_j(r_j)(1 - w(N \setminus \{j\}))$ . Hence we have:  $(u_i(y_i) - u_i(x_i))v(\{i\}) = (u_j(R_j) - u_j(r_j))(1 - w(N \setminus \{j\}))$  (c).

Similarly, for any  $h_i$  element of  $X_i \setminus \{x_i\}$  chosen such that  $h_i \succ_i \mathbf{0}_i$ , i.e.,  $u_i(h_i) \geq 0$ , if  $z_i$  is the answer to question  $Q_j^+(h_i)$  we have:  $(u_i(z_i) - u_i(h_i))v(\{i\}) = (u_j(R_j) - u_j(r_j))(1 - w(N \setminus \{j\}))$  (d).

From Equations (c) and (d) we have  $(u_i(y_i) - u_i(x_i))v(\{i\}) = (u_i(z_i) - u_i(h_i))v(\{i\})$ . Assuming  $(\mathbf{1}_i, \mathbf{0}_{-i}) \succ \mathbf{0}$ , i.e.,  $v(\{i\}) > 0$  we obtain:

$$u_i(y_i) - u_i(x_i) = u_i(z_i) - u_i(h_i)$$

By taking  $h_i = y_i$ , we can obtain the following simplified equation:

$$u_i(y_i) - u_i(x_i) = u_i(z_i) - u_i(y_i)$$

### 8.3 Application to preference learning over sets

Recall that in the context of sets all  $X_i$  are discrete and equal to the set of items  $E$ . Moreover all  $u_i$  are identical and denoted  $u$  hereafter.

#### Utility elicitation below the neutral level

Let  $i, j \in \{1, \dots, n\}$ ,  $x_i, r_j, R_j \in E$  such that  $x_i \succ \mathbf{0}$  and  $R_j \succ r_j \succ \mathbf{0}$ . We consider the following query:

$LQ_j^-(x_i)$  : what is the element  $y_i^+$  of lowest index such that  $(x_i, r_j, \mathbf{0}_{-ij}) \succ (y_i^+, R_j, \mathbf{0}_{-ij})$ ?

From this preference we deduce that  $(x_i, r_j, \mathbf{0}_{-ij}) \succ (y_i^-, R_j, \mathbf{0}_{-ij})$  where  $y_i^+ = e_{k+1}, y_i^- = e_k$  for some  $k \in \{1, \dots, |E| - 1\}$ . Hence we obtain the following inequations:  $(u(x_i) - u(y_i^+))(1 - w(N \setminus \{i\})) \leq (u(R_j) - u(r_j))v(\{j\})$  and  $(u(x_i) - u(y_i^-))(1 - w(N \setminus \{i\})) > (u(R_j) - u(r_j))v(\{j\})$ .

Similarly, for any  $h_i$  element of  $E \setminus \{x_i\}$  chosen such that  $h_i \succ \mathbf{0}$ , we ask the following question:

$HQ_j^-(h_i)$  : what is the element  $z_i^-$  of highest index such that  $(h_i, r_j, \mathbf{0}_{-ij}) \succ (z_i^-, R_j, \mathbf{0}_{-ij})$ ?

From this preference we deduce that  $(h_i, r_j, \mathbf{0}_{-ij}) \prec (z_i^+, R_j, \mathbf{0}_{-ij})$ . Hence we obtain the following inequations:  $(u(h_i) - u(z_i^-))(1 - w(N \setminus \{i\})) \geq (u(R_j) - u(r_j))v(\{j\})$  and  $(u(h_i) - u(z_i^+))(1 - w(N \setminus \{i\})) < (u(R_j) - u(r_j))v(\{j\})$ .

Hence we have  $(u(x_i) - u(y_i^+))(1 - w(N \setminus \{i\})) \leq (u(R_j) - u(r_j))v(\{j\}) \leq (u(h_i) - u(z_i^-))(1 - w(N \setminus \{i\}))$ . Moreover,  $(u(h_i) - u(z_i^+))(1 - w(N \setminus \{i\})) > (u(R_j) - u(r_j))v(\{j\}) > (u(x_i) - u(y_i^-))(1 - w(N \setminus \{i\}))$ . Assuming  $(-\mathbf{1}_i, \mathbf{0}_{-i}) \prec \mathbf{0}$ , i.e.,  $w(N \setminus \{i\}) < 1$ , we obtain:

$$u(h_i) - u(z_i^-) \geq u(x_i) - u(y_i^+)$$

$$u(h_i) - u(z_i^+) > u(x_i) - u(y_i^-)$$

By taking  $h_i = y_i^+$ , we can obtain the following simplified inequations:

$$u(y_i^+) - u(z_i^-) \geq u(x_i) - u(y_i^+)$$

$$u(y_i^+) - u(z_i^+) > u(x_i) - u(y_i^-)$$

#### Utility elicitation above the neutral level

Let  $i, j \in \{1, \dots, n\}$ ,  $x_i, r_j, R_j \in E$  such that  $x_i \succ \mathbf{0}$  and  $r_j \prec R_j \prec \mathbf{0}$ . We consider the following query:

$HQ_j^+(x_i)$  : what is the element  $y_i^-$  of highest index such that  $(x_i, R_j, \mathbf{0}_{-ij}) \succ (y_i^-, r_j, \mathbf{0}_{-ij})$ ?

From this preference we deduce that  $(x_i, R_j, \mathbf{0}_{-ij}) \prec (y_i^+, r_j, \mathbf{0}_{-ij})$  where  $y_i^+ = e_{k+1}, y_i^- = e_k$  for some  $k \in \{1, \dots, |E| - 1\}$ . Hence we obtain the following inequations:  $(u(x_i) - u(y_i^-))v(\{i\}) \geq (u(r_j) - u(R_j))(1 - w(N \setminus \{j\}))$  and  $(u(x_i) - u(y_i^+))v(\{i\}) < (u(r_j) - u(R_j))(1 - w(N \setminus \{j\}))$ .

Similarly, for any  $h_i$  element of  $E \setminus \{x_i\}$  chosen such that  $h_i \succ \mathbf{0}$ , we ask the following question:

$LQ_j^+(h_i)$  : what is the element  $z_i^+$  of lowest index such that  $(h_i, R_j, \mathbf{0}_{-ij}) \preceq (z_i^+, r_j, \mathbf{0}_{-ij})$ ?

From this preference we deduce that  $(h_i, R_j, \mathbf{0}_{-ij}) \succ (z_i^-, r_j, \mathbf{0}_{-ij})$  where  $z_i^+ = e_{k+1}, z_i^- = e_k$  for some  $k \in \{1, \dots, |E| - 1\}$ . Hence we obtain the following inequations:  $(u(h_i) - u(z_i^+))v(\{i\}) \leq (u(r_j) - u(R_j))(1 - w(N \setminus \{j\}))$  and  $(u(h_i) - u(z_i^-))v(\{i\}) > (u(r_j) - u(R_j))(1 - w(N \setminus \{j\}))$ .

Hence we have  $(u(h_i) - u(z_i^+))v(\{i\}) \leq (u(R_j) - u(r_j))(1 - w(N \setminus \{j\})) \leq (u(x_i) - u(y_i^-))v(\{i\})$ . Moreover,  $(u(x_i) - u(y_i^+))v(\{i\}) < (u(R_j) - u(r_j))(1 - w(N \setminus \{j\})) < (u(h_i) - u(z_i^-))v(\{i\})$ . Assuming  $(\mathbf{1}_i, \mathbf{0}_{-i}) \succ \mathbf{0}$ , i.e.,  $v(\{i\}) > 0$ , we obtain:

$$\begin{aligned} u(h_i) - u(z_i^+) &\leq u(x_i) - u(y_i^-) \\ u(h_i) - u(z_i^-) &> u(x_i) - u(y_i^+) \end{aligned}$$

By taking  $h_i = y_i^-$ , we can obtain the following simplified inequations:

$$\begin{aligned} u(y_i^-) - u(z_i^+) &\leq u(x_i) - u(y_i^-) \\ u(y_i^-) - u(z_i^-) &> u(x_i) - u(y_i^+) \end{aligned}$$