



HAL
open science

The multilinear model in multicriteria decision making: The case of 2-additive capacities and contributions to parameter identification

Guilherme Dean Pelegrina, Leonardo Tomazeli Duarte, Michel Grabisch, João
Marcos Travassos Romano

► To cite this version:

Guilherme Dean Pelegrina, Leonardo Tomazeli Duarte, Michel Grabisch, João Marcos Travassos Romano. The multilinear model in multicriteria decision making: The case of 2-additive capacities and contributions to parameter identification. *European Journal of Operational Research*, 2020, 282, pp.945-956. 10.1016/j.ejor.2019.10.005 . halshs-02379646

HAL Id: halshs-02379646

<https://shs.hal.science/halshs-02379646>

Submitted on 25 Nov 2019

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.

The multilinear model in multicriteria decision making: The case of 2-additive capacities and contributions to parameter identification

Guilherme Dean Pelegrina^{a,b,*}, Leonardo Tomazeli Duarte^c, Michel Grabisch^{b,d}, João Marcos Travassos Romano^a

^a*School of Electrical and Computer Engineering, University of Campinas, 400 Albert Einstein Avenue, 13083-852 Campinas, Brazil*

^b*Centre d'Économie de la Sorbonne, Université Paris I - Panthéon-Sorbonne, 106-112 Boulevard de l'Hôpital, 75647 Paris Cedex 13, France*

^c*School of Applied Sciences, University of Campinas, 1300 Pedro Zaccaria Street, 13484-350 Limeira, Brazil*

^d*Paris School of Economics, Université Paris I – Panthéon-Sorbonne, 106-112 Boulevard de l'Hôpital, 75647 Paris Cedex 13, France*

Abstract

In several multicriteria decision making problems, it is important to consider interactions among criteria in order to satisfy the preference relations provided by the decision maker. This can be achieved by using aggregation functions based on fuzzy measures, such as the Choquet integral and the multilinear model. Although the Choquet integral has been studied in a large number of works, one does not find the same literature with respect to the multilinear model. In this context, the contribution of this work is twofold. We first provide a formulation of the multilinear model by means of a 2-additive capacity. A second contribution lies in the problem of capacity identification. We consider a supervised approach and apply optimization models with and without regularization terms. Results obtained in numerical experiments with both synthetic and real data attest the performance of the considered approaches.

Keywords: multiple criteria analysis, multi-attribute utility theory, multilinear model, 2-additive capacity, capacity identification

1. Introduction

Consider a set of objects of interest (or alternatives) $X \subseteq X_1 \times \dots \times X_m$ evaluated according to a set C of m criteria. Let us assume that the evaluation of an alternative $\mathbf{x} \in X$ with respect to the criterion i is given by the value function $u_i(x_i)$, $x_i \in X_i$. Moreover, consider that the global

*Corresponding author

Email addresses: pelegrina@decom.fee.unicamp.br (Guilherme Dean Pelegrina), leonardo.duarte@fca.unicamp.br (Leonardo Tomazeli Duarte), michel.grabisch@univ-paris1.fr (Michel Grabisch), romano@dm.o.fee.unicamp.br (João Marcos Travassos Romano)

evaluation of \mathbf{x} is given by the overall value function

$$U(\mathbf{x}) = F(u(\mathbf{x})), \quad (1)$$

where $u(\mathbf{x}) = (u_1(x_1), \dots, u_m(x_m))$ and $F(\cdot)$ is an aggregation function, which is nondecreasing in its arguments. This model, called decomposable or separable (see (Blackorby et al., 1978)), has been widely studied in the literature (see, e.g., the works of Greco et al. (2004) and Bouyssou and Pirlot (2004)). For simplicity of notation, let us refer to $u_i(x_i)$ as u_i .

In multi-attribute utility theory (MAUT) (Keeney and Raiffa, 1976), the decision maker is generally concerned in how to model preference relations among the set of alternatives, represented by

$$\mathbf{x} \succeq \mathbf{x}' \Leftrightarrow U(\mathbf{x}) \geq U(\mathbf{x}'), \quad \forall \mathbf{x}, \mathbf{x}' \in X, \quad (2)$$

where \succeq indicates that \mathbf{x} is at least as good as \mathbf{x}' . Therefore, the preference relations (2) depend on the form of the aggregation function $F(\cdot)$, which is based on the hypotheses about the addressed multicriteria decision making (MCDM) problem (Figueira et al., 2016).

A well-known aggregation function is the Weighted Arithmetic Mean (WAM), expressed by

$$F_{WAM}(u(\mathbf{x})) = \sum_{i=1}^m w_i u_i, \quad (3)$$

where w_i ($w_i \geq 0$, $\sum_{i=1}^m w_i = 1$) represents the weight factor associated with criterion i . In order to apply the WAM, one assumes that the decision criteria are mutually preferentially independent (Keeney and Raiffa, 1976). However, since this assumption does not hold in several practical situations, there are some limitations by using the WAM to model preference relations (Grabisch and Labreuche, 2016). In this context, one may use an aggregation function that takes into account the interaction among criteria, such as the Choquet integral (Choquet, 1954) or the multilinear model (Owen, 1972). For instance, Choquet integral has been applied in several areas, such as in ergonomics (Raufaste et al., 2001; Grabisch et al., 2006), supply chain management (Feyzioglu and Büyüközkan, 2010), tourism (Li et al., 2013) and project evaluations (Bottero et al., 2018).

Both the Choquet integral and the multilinear model represent the interaction among criteria through a capacity (Choquet, 1954), which is a set function $\mu : 2^C \rightarrow \Re$ satisfying the following axioms:

- $\mu(\emptyset) = 0$ and $\mu(C) = 1$ (boundedness),

- if $A \subseteq B \subseteq C$, $\mu(A) \leq \mu(B) \leq \mu(C)$ (monotonicity).

The (discrete) Choquet integral (Choquet, 1954) takes into account the interaction among criteria through a piecewise linear procedure. It is defined as follows:

$$F_{CI}(\mathbf{x}) = \sum_{i=1}^m (u_{(i)} - u_{(i-1)}) \mu(\{(i), \dots, (m)\}), \quad (4)$$

where $u_{(i)}$ indicates a permutation of the indices i so that $0 \leq u_{(1)} \leq \dots \leq u_{(m)} \leq 1$ (with $u_{(0)} = 0$). An interesting remark is that, assuming that all the value functions $u_i(\cdot)$, $i = 1, \dots, m$, are commensurate, the Choquet integral is obtained as the parsimonious linear interpolation of $F(\cdot)$ over the vertices of the hypercube $[0, 1]^m$ (Grabisch, 2016). Therefore, the capacity coefficients $\mu(A)$ that are used in (4) depend on the order that we have in the set of evaluations (u_1, \dots, u_m) .

A second popular nonlinear aggregation function is the multilinear model (Owen, 1972; Keeney and Raiffa, 1976), which is defined as follows:

$$F_{ML}(\mathbf{x}) = \sum_{A \subseteq C} \mu(A) \prod_{i \in A} u_i \prod_{i \in \bar{A}} (1 - u_i), \quad (5)$$

where $u_i \in [0, 1]$ and \bar{A} is the complement set of A . It is an approach that comprises a polynomial aggregation of the criteria evaluations. Moreover, differently from the Choquet integral, (5) does not need the commensurability assumption and is obtained as the linear interpolation using all vertices of the hypercube $[0, 1]^m$ (Grabisch, 2016). Therefore, in the multilinear model, all the capacity coefficients are used to aggregate the criteria evaluations.

One may note that, in both cases, other than the parameters associated with each individual criterion, one also considers the ones that are associated with coalitions of two or more criteria. In this case, the number of parameters to be determined, given by $2^m - 2$, increases exponentially with the number of criteria, which may bring a difficulty for the decision maker to perform capacity identification. Therefore, in order to deal with this issue, one may use techniques based on learning through data (the focus of this paper), semantics or a combination of both (Grabisch, 1996). Moreover, one may also consider a 2-additive capacity (Grabisch, 1997), which simplifies the model by taking into account only interactions among pairs of criteria and the parameters associated with singletons. In this case, one reduces the number of parameters to be identified to $m(m+1)/2 - 1$. For further discussions about 2-additive aggregation functions see Kolesárová et al. (2018).

Several works in the literature address the capacity identification problem in the context of Cho-

quet integral (Grabisch et al., 2008). For instance, one may find supervised approaches (Grabisch, 1995; Marichal and Roubens, 2000; Miranda et al., 2003; Combarro and Miranda, 2006; Angilella et al., 2010, 2015), supervised approaches with regularization (Anderson et al., 2014; Adeyeba et al., 2015; Oliveira et al., 2017) and non-supervised approaches (Kojadinovic, 2004, 2008; Rowley et al., 2015; Duarte, 2018). However, one does not find the same references with respect to the multilinear model. In view of this lack of literature, the contribution of this paper is twofold¹. The first one is to exploit the concept of 2-additive capacity in the context of the multilinear model. The obtained expression, which is, as far as we know, unknown in the literature, has some similarities with respect to the one in the context of Choquet integral. Therefore, we also provide a parallel between these two expressions. As a second contribution, we address the problem of capacity identification by means of a supervised approach. For instance, based on learning data consisting of a set of evaluations with respect to a set of alternatives and their associated overall values, our aim here is to find the multilinear model parameters that can restore as well as possible the given overall values. Moreover, we also exploit the application of a regularization term in the optimization model, which may contribute to find a capacity that can better deal with new data other than the one used in the learning algorithm.

The rest of this paper is organized as follows. Section 2 presents the concept of interaction among criteria. In Section 3, we formalize the multilinear model expression by using a 2-additive capacity and provide a comparison with the Choquet integral. Section 4 addresses the problem of capacity identification. In Section 5, we conduct numerical experiments in both synthetic and real data. Finally, our conclusions and future perspectives are described in Section 6.

2. The notion of interaction among criteria

A desired feature in MCDM methods is how to interpret the parameters of the considered aggregation function. In the context of the multilinear model, it has been shown that (see Grabisch et al. (2000) and Grabisch (2016), Ch. 6) this can be achieved by using the Banzhaf interaction index proposed by Roubens (1996), which is defined by

$$I^B(A) = \frac{1}{2^{|C|-|A|}} \sum_{D \subseteq C \setminus A} \sum_{D' \subseteq A} (-1)^{|A|-|D'|} \mu(D' \cup D), \forall A \subseteq C, \quad (6)$$

¹It is worth mentioning that a preliminary discussion on this subject was presented in the conference paper (Pelegrina et al., 2018).

where $|C|$ and $|A|$ represent the cardinalities of the set C and subset A , respectively. It is worth mentioning that this concept is based on the work developed by Banzhaf in game theory (Banzhaf, 1965). Moreover, one may note that, given $I^{\mathcal{B}}(A)$, $\forall A \subseteq C$, the capacity can be retrieved through the equation

$$\mu(A) = \sum_{D \subseteq C} \left(\frac{1}{2}\right)^{|D|} (-1)^{|D \setminus A|} I^{\mathcal{B}}(D), \forall A \subseteq C. \quad (7)$$

For coalitions of two or more criteria, $I^{\mathcal{B}}(A)$ is difficult to interpret. However, if $|A| \leq 2$, the obtained values have a clear meaning. For instance, if one takes into account a singleton i , (6) leads to the Banzhaf power index $\phi^{\mathcal{B}}(\{i\})$, expressed by

$$\phi^{\mathcal{B}}(\{i\}) = \frac{1}{2^{|C|-1}} \sum_{D \subseteq C \setminus \{i\}} [\mu(D \cup \{i\}) - \mu(D)]. \quad (8)$$

This index, which lies in the range $[0, 1]$, can be interpreted as the marginal contribution of criterion i alone in all coalitions. Now, if one considers a pair of criteria i, i' , (6) leads to

$$I^{\mathcal{B}}(\{i, i'\}) = \frac{1}{2^{|C|-2}} \sum_{D \subseteq C \setminus \{i, i'\}} [\mu(D \cup \{i, i'\}) - \mu(D \cup \{i\}) - \mu(D \cup \{i'\}) + \mu(D)]. \quad (9)$$

In this case, $I^{\mathcal{B}}(\{i, i'\})$ lies in the range $[-1, 1]$ and can be viewed as the interaction degree of coalition of criteria i, i' , taking into account all possible coalitions. Moreover, $I^{\mathcal{B}}(\{i, i'\}) < 0$ indicates a negative interaction (or a redundant effect) among criteria i, i' , which means that it is sufficient to consider either i or i' in order to achieve a good global evaluation. On the other hand, $I^{\mathcal{B}}(\{i, i'\}) > 0$ indicates a positive interaction (or a complementary effect) among criteria i, i' , which means that it is important to satisfy both i and i' in order to achieve a good global evaluation. $I^{\mathcal{B}}(\{i, i'\}) = 0$ indicates no interaction and the criteria i, i' act independently (Grabisch, 2000). From now on, in order to avoid a heavy notation, we refer to $\phi^{\mathcal{B}}(\{i\})$ and $I^{\mathcal{B}}(\{i, i'\})$ as $\phi_i^{\mathcal{B}}$ and $I_{i, i'}^{\mathcal{B}}$, respectively.

3. The 2-additive multilinear model

In this section, we derive the expression of the multilinear model when the underlying capacity is 2-additive. For instance, consider the object of interest $\mathbf{x} \in X$ and its evaluations $u(\mathbf{x}) = (u_1, u_2, \dots, u_m)$. An alternative representation of the multilinear model is given by

$$F_{ML}(\mathbf{x}) = \sum_{A \subseteq C} a(A) \prod_{i \in A} u_i, \quad (10)$$

where $a(A)$ is the Möbius transform of μ (Rota, 1964), given by

$$a(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \mu(B), \forall A \subseteq C. \quad (11)$$

One may also express the Möbius transform in terms of $I^{\mathcal{B}}(B)$, i.e.,

$$a(A) = \sum_{B \supseteq A} \left(-\frac{1}{2}\right)^{|B|-|A|} I^{\mathcal{B}}(B), \forall A \subseteq C. \quad (12)$$

Moreover, given the Möbius transform, the capacity can be retrieved as follows:

$$\mu(A) = \sum_{D \subseteq A} a(D), \forall A \subseteq C. \quad (13)$$

A 2-additive capacity is such that $a(A) = 0$ for any $A \subseteq C$, $|A| > 2$ and $I^{\mathcal{B}}(B) = 0$ for any $B \subseteq C$, $|B| > 2$ (Grabisch, 1997). Therefore,

$$\begin{aligned} F_{ML}(\mathbf{x}) &= \sum_{\substack{A \subseteq C, \\ |A| \leq 2}} a(A) \prod_{i \in A} u_i \\ &= \sum_{\substack{A \subseteq C, \\ |A| \leq 2}} \sum_{\substack{B \supseteq A, \\ |B| \leq 2}} \left(-\frac{1}{2}\right)^{|B|-|A|} I^{\mathcal{B}}(B) \prod_{i \in A} u_i \\ &= \sum_{\substack{B, \\ |B| \leq 2}} \left(-\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) + \sum_i \sum_{\substack{B \ni i, \\ |B| \leq 2}} \left(-\frac{1}{2}\right)^{|B|-1} I^{\mathcal{B}}(B) u_i + \sum_{\{i, i'\}} \sum_{\substack{B \supseteq \{i, i'\}, \\ |B| \leq 2}} \left(-\frac{1}{2}\right)^{|B|-2} I^{\mathcal{B}}(B) u_i u_{i'} \\ &= I^{\mathcal{B}}(\emptyset) - \frac{1}{2} \sum_i \phi_i^{\mathcal{B}} + \frac{1}{4} \sum_{i, i'} I_{i, i'}^{\mathcal{B}} + \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i'} I_{i, i'}^{\mathcal{B}} \right) + \sum_{i, i'} u_i u_{i'} I_{i, i'}^{\mathcal{B}} \end{aligned}$$

If we express the axioms of a capacity in a 2-additive model and in terms of the Banzhaf interaction index, one achieves the following (see Appendix A for further details):

- $\mu(\emptyset) = 0 \rightarrow I^{\mathcal{B}}(\emptyset) - \frac{1}{2} \sum_i \phi_i^{\mathcal{B}} + \frac{1}{4} \sum_{i, i'} I_{i, i'}^{\mathcal{B}} = 0;$
- $\mu(C) = 1 \rightarrow I^{\mathcal{B}}(\emptyset) + \frac{1}{2} \sum_i \phi_i^{\mathcal{B}} + \frac{1}{4} \sum_{i, i'} I_{i, i'}^{\mathcal{B}} = 1;$
- $\mu(\{A \cup i\}) - \mu(A) \geq 0 \rightarrow \phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i' \neq i} \left| I_{i, i'}^{\mathcal{B}} \right| \geq 0, \forall i \in C.$

Moreover, since $I_{i,i'}^{\mathcal{B}}$ may be positive or negative, we can write for every $i \in C$

$$\sum_{i'} I_{i,i'}^{\mathcal{B}} = \sum_{i'} |I_{i,i'}^{\mathcal{B}}| - 2 \sum_{\substack{i', \\ I_{i,i'}^{\mathcal{B}} < 0}} |I_{i,i'}^{\mathcal{B}}|.$$

Therefore, based on the aforementioned results, one obtains

$$\begin{aligned} F_{ML}(\mathbf{x}) &= \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i'} I_{i,i'}^{\mathcal{B}} \right) + \sum_{i,i'} u_i u_{i'} I_{i,i'}^{\mathcal{B}} \\ F_{ML}(\mathbf{x}) &= \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i'} |I_{i,i'}^{\mathcal{B}}| \right) + \sum_i u_i \sum_{i', I_{i,i'}^{\mathcal{B}} < 0} |I_{i,i'}^{\mathcal{B}}| + \sum_{i,i'} u_i u_{i'} I_{i,i'}^{\mathcal{B}} \\ &= \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i'} |I_{i,i'}^{\mathcal{B}}| \right) + \sum_{I_{i,i'}^{\mathcal{B}} < 0} (u_i + u_{i'}) |I_{i,i'}^{\mathcal{B}}| + \sum_{I_{i,i'}^{\mathcal{B}} > 0} u_i u_{i'} I_{i,i'}^{\mathcal{B}} - \sum_{I_{i,i'}^{\mathcal{B}} < 0} u_i u_{i'} |I_{i,i'}^{\mathcal{B}}| \\ &= \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i'} |I_{i,i'}^{\mathcal{B}}| \right) + \sum_{I_{i,i'}^{\mathcal{B}} < 0} (u_i + u_{i'} - u_i u_{i'}) |I_{i,i'}^{\mathcal{B}}| + \sum_{I_{i,i'}^{\mathcal{B}} > 0} u_i u_{i'} I_{i,i'}^{\mathcal{B}} \end{aligned}$$

Finally, we observe that the above expression comprises an additive term, a disjunctive term $\mathbf{N}_{\mathbf{p}}(u_i, u_{i'}) = u_i + u_{i'} - u_i u_{i'}$, which turns out to be the t-conorm usually called probabilistic sum of $(u_i, u_{i'})$, and a conjunctive term $\mathbf{T}_{\mathbf{p}}(u_i, u_{i'}) = u_i u_{i'}$, which is the product t-norm of $(u_i, u_{i'})$ (Klement et al., 2000; Beliakov et al., 2007). Therefore, the 2-additive multilinear model may be written as

$$F_{ML}(\mathbf{x}) = \sum_i u_i \left(\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i'} |I_{i,i'}^{\mathcal{B}}| \right) + \sum_{I_{i,i'}^{\mathcal{B}} < 0} \mathbf{N}_{\mathbf{p}}(u_i, u_{i'}) |I_{i,i'}^{\mathcal{B}}| + \sum_{I_{i,i'}^{\mathcal{B}} > 0} \mathbf{T}_{\mathbf{p}}(u_i, u_{i'}) I_{i,i'}^{\mathcal{B}}. \quad (14)$$

We may remark that (14) is similar to the 2-additive expression of the Choquet integral (Grabisch, 2000), which is given by

$$F_{CI}(\mathbf{x}) = \sum_i u_i \left(\phi_i^{\mathcal{S}} - \frac{1}{2} \sum_{i'} |I_{i,i'}^{\mathcal{S}}| \right) + \sum_{I_{i,i'}^{\mathcal{S}} < 0} (u_i \vee u_{i'}) |I_{i,i'}^{\mathcal{S}}| + \sum_{I_{i,i'}^{\mathcal{S}} > 0} (u_i \wedge u_{i'}) I_{i,i'}^{\mathcal{S}}, \quad (15)$$

where $\phi_i^{\mathcal{S}}$ and $I_{i,i'}^{\mathcal{S}}$ are the power and the interaction indices corresponding to the Shapley value (Shapley, 1953; Grabisch, 1997), \vee represents the maximum operator (also a t-conorm) and \wedge represents the minimum operator (also a t-norm). In that respect, since for a 2-additive capacity $\phi_i^{\mathcal{B}} = \phi_i^{\mathcal{S}} =: \phi_i$ and $I_{i,i'}^{\mathcal{B}} = I_{i,i'}^{\mathcal{S}} =: I_{i,i'}$ (Marichal and Roubens, 2000), one may generalize both aggregation functions

by the following:

$$F_{\mathbf{N},\mathbf{T}}(\mathbf{x}) = \sum_i u_i \left(\phi_i - \frac{1}{2} \sum_{i'} |I_{i,i'}| \right) + \sum_{I_{i,i'} < 0} \mathbf{N}(u_i, u_{i'}) |I_{i,i'}| + \sum_{I_{i,i'} > 0} \mathbf{T}(u_i, u_{i'}) I_{i,i'}, \quad (16)$$

where \mathbf{N} is a t-conorm and \mathbf{T} is a t-norm².

3.1. A graphical interpretation when $m = 2$

We here provide a graphical interpretation of the aggregation of the 2-additive multilinear model. We consider the same approach discussed in (Grabisch, 2000) and we restrict our analysis to the case of $m = 2$. In this case, one may represent the multilinear model by

$$F_{ML}(\mathbf{x}) = \begin{cases} u_1 (\phi_1^{\mathcal{B}} - \frac{1}{2} I_{1,2}^{\mathcal{B}}) + u_2 (\phi_2^{\mathcal{B}} - \frac{1}{2} I_{1,2}^{\mathcal{B}}) + \mathbf{T}_{\mathbf{p}}(u_1, u_2) I_{1,2}^{\mathcal{B}}, & I_{1,2}^{\mathcal{B}} \geq 0 \\ u_1 (\phi_1^{\mathcal{B}} + \frac{1}{2} I_{1,2}^{\mathcal{B}}) + u_2 (\phi_2^{\mathcal{B}} + \frac{1}{2} I_{1,2}^{\mathcal{B}}) + \mathbf{N}_{\mathbf{p}}(u_1, u_2) |I_{1,2}^{\mathcal{B}}|, & I_{1,2}^{\mathcal{B}} \leq 0 \end{cases} \quad (17)$$

Moreover, by exploiting the axioms of a capacity, it is possible to obtain the following:

- $I^{\mathcal{B}}(\emptyset) - \frac{1}{2} \sum_i \phi_i^{\mathcal{B}} + \frac{1}{4} \sum_{i,i'} I_{i,i'}^{\mathcal{B}} = 0 \rightarrow I^{\mathcal{B}}(\emptyset) - \frac{1}{2} (\phi_1^{\mathcal{B}} + \phi_2^{\mathcal{B}}) + \frac{1}{4} I_{1,2}^{\mathcal{B}} = 0;$
- $I^{\mathcal{B}}(\emptyset) + \frac{1}{2} \sum_i \phi_i^{\mathcal{B}} + \frac{1}{4} \sum_{i,i'} I_{i,i'}^{\mathcal{B}} = 1 \rightarrow I^{\mathcal{B}}(\emptyset) + \frac{1}{2} (\phi_1^{\mathcal{B}} + \phi_2^{\mathcal{B}}) + \frac{1}{4} I_{1,2}^{\mathcal{B}} = 1;$
- $\phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i' \neq i} |I_{i,i'}^{\mathcal{B}}| \geq 0 \rightarrow \phi_i^{\mathcal{B}} \pm \frac{1}{2} I_{1,2}^{\mathcal{B}} \geq 0, \forall i \in \{1, 2\}.$

Based on the first two results, one concludes that $\phi_1^{\mathcal{B}} + \phi_2^{\mathcal{B}} = 1$. Therefore, one may represent this scenario in the $(\phi_1^{\mathcal{B}}, I_{1,2}^{\mathcal{B}})$ coordinates, which is illustrated in Figure 1.

If we consider the horizontal axis, in which $I_{1,2}^{\mathcal{B}} = 0$, Equation (17) becomes

$$F_{ML}(\mathbf{x}) = u_1 \phi_1^{\mathcal{B}} + u_2 \phi_2^{\mathcal{B}}, \quad (18)$$

i.e. a weighted mean. One may remark that, for the extreme cases of $\phi_1^{\mathcal{B}} = 0$ or $\phi_2^{\mathcal{B}} = 0$, one obtains $F_{ML}(\mathbf{x}) = u_2$ or $F_{ML}(\mathbf{x}) = u_1$, respectively.

On the other hand, if we consider the vertical axis, in which $\phi_1^{\mathcal{B}} = \phi_2^{\mathcal{B}} = 1/2$, Equation (17) becomes

$$F_{ML}(\mathbf{x}) = u_1 \left(\frac{1}{2} - \frac{1}{2} I_{1,2}^{\mathcal{B}} \right) + u_2 \left(\frac{1}{2} - \frac{1}{2} I_{1,2}^{\mathcal{B}} \right) + u_1 u_2 I_{1,2}^{\mathcal{B}}, \quad (19)$$

²For further discussion on generalizations of the Choquet integral and the multilinear model, see (Kolesárová et al., 2012), who argue that \mathbf{T} and \mathbf{N} should be rather a (quasi-)copula and its dual $\mathbf{N}(x, y) = x + y - \mathbf{T}(x, y)$.

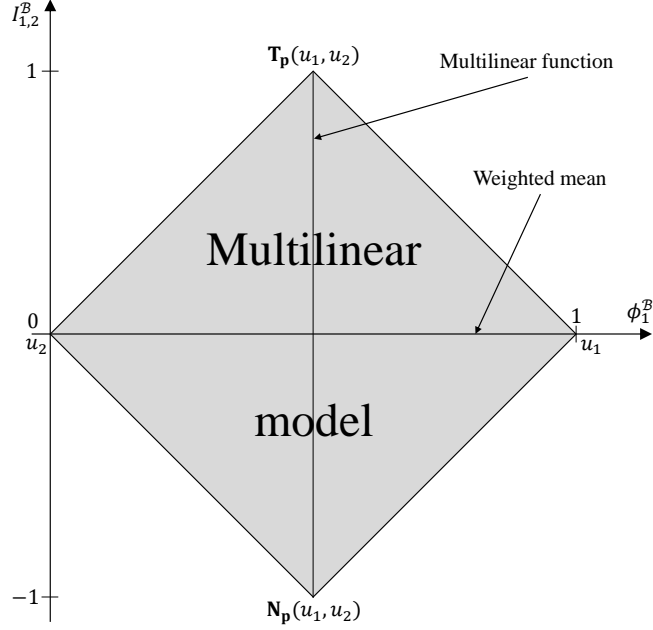


Figure 1: Graphical interpretation of multilinear model when $m = 2$ criteria.

i.e., a bilinear function (Keeney and Raiffa, 1976) with weights $(\frac{1}{2} - \frac{1}{2}I_{1,2}^B)$, $(\frac{1}{2} - \frac{1}{2}I_{1,2}^B)$ and $I_{1,2}^B$. One also may remark that, for the extreme cases of $I_{1,2}^B = 1$ or $I_{1,2}^B = -1$, one obtains $F_{ML}(\mathbf{x}) = \mathbf{T}_p(u_1, u_2) = u_1u_2$ or $F_{ML}(\mathbf{x}) = \mathbf{N}_p(u_1, u_2) = u_1 + u_2 - u_1u_2$, respectively.

4. The problem of capacity identification

An important task in MAUT is how to define the parameters of the aggregation function. In this section, we address the problem of capacity identification in the context of the multilinear model by means of a supervised approach, i.e., based on learning data. For instance, we consider as learning data the evaluations $u_{i,j}$ of a set of n alternatives $\mathbf{x}_j \in X$, $j = 1, \dots, n$, and their associated overall values $y(u(\mathbf{x}_j))$. These data will be used to retrieve a capacity μ that leads to overall evaluations $F_{ML}(u(\mathbf{x}_j))$ as close as possible to $y(u(\mathbf{x}_j))$, for all $j = 1, \dots, n$. As will be discussed in the sequel, one exploits identification methods with and without the application of regularization terms.

4.1. Supervised approach

Consider a supervised approach whose aim is to minimize the mean squared error between the obtained evaluations $F_{ML}(u(\mathbf{x}_j))$ and the desired ones $y(u(\mathbf{x}_j))$ (collected from learning data).

Mathematically, this cost function (or representation error) is given by

$$E = \sum_{j=1}^n (F_{ML}(u(\mathbf{x}_j)) - y(u(\mathbf{x}_j)))^2. \quad (20)$$

Since the aggregation in multilinear model is linear with respect to the capacity coefficients, one can also represent (20) in terms of vectors and matrices. In this case, consider the set of capacity values $\mu = [\mu(\emptyset), \mu(\{1\}), \dots, \mu(\{1, 2\}), \dots, \mu(\{1, 2, \dots\}), \dots, \mu(C)]^T$ in cardinal-lexicographic representation³, the set of desired evaluation $\mathbf{y} = [y(u(\mathbf{x}_1)), y(u(\mathbf{x}_2)), \dots, y(u(\mathbf{x}_n))]^T$ and the following matrix:

$$\mathbf{P} = \begin{bmatrix} \prod_{i \in C} (1 - u_{i,1}) & \prod_{i \in C} (1 - u_{i,2}) & \dots & \prod_{i \in C} (1 - u_{i,n}) \\ u_{1,1} \prod_{i \in \overline{\{1\}}} (1 - u_{i,1}) & u_{1,2} \prod_{i \in \overline{\{1\}}} (1 - u_{i,2}) & \dots & u_{1,n} \prod_{i \in \overline{\{1\}}} (1 - u_{i,n}) \\ \vdots & \vdots & \ddots & \vdots \\ \prod_{i \in C} u_{i,1} & \prod_{i \in C} u_{i,2} & \dots & \prod_{i \in C} u_{i,n} \end{bmatrix}.$$

Since the minimization of $E = \sum_{j=1}^n (F_{ML}(u(\mathbf{x}_j)) - y(u(\mathbf{x}_j)))^2 = \mu^T \mathbf{P} \mathbf{P}^T \mu - 2\mathbf{y}^T \mathbf{P}^T \mu + \mathbf{y}^T \mathbf{y}$ is equivalent to the minimization of $\mu^T \mathbf{P} \mathbf{P}^T \mu - 2\mathbf{y}^T \mathbf{P}^T \mu$, the optimization problem can be represented in a quadratic form by

$$\begin{aligned} \min_{\mu} \quad & \frac{1}{2} \mu^T \mathbf{Q} \mu + \mathbf{v}^T \mu \\ \text{s.t.} \quad & \mathbf{L} \mu = [0, 1]^T \\ & \mathbf{M} \mu \leq \mathbf{0} \end{aligned} \quad (21)$$

where $\mathbf{Q} = 2\mathbf{P} \mathbf{P}^T$, $\mathbf{v} = -2\mathbf{P} \mathbf{y}$ and the matrices \mathbf{L} and \mathbf{M} guarantee that the axioms of a capacity (boundedness and monotonicity, respectively) are satisfied. For instance, in a scenario with $m = 2$ criteria, we have

$$\mathbf{L} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathbf{M} = \begin{bmatrix} 1 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

³We refer to μ in cardinal-lexicographic representation as a vector in which the elements are sorted according to their cardinality and, for each cardinality, based on the lexicographic order.

4.1.1. Example

In this section, we provide an illustrative example in order to verify the performance of the quadratic model (21) in several synthetic decision problems. Other than the multilinear model, we also considered the identification problem with respect to the Choquet integral⁴. For instance, we consider decision problems with $m = 3$, $m = 4$ and $m = 5$ criteria and with a number of alternatives varying from 1 to 50. With respect to the performance index, we calculate the squared error

$$e_\mu = \sum_{A \subseteq C, A \neq \emptyset, A \neq C} (\mu(A) - \hat{\mu}(A))^2, \quad (22)$$

where μ is the capacity used to obtain the global evaluations $y(u(\mathbf{x}_j))$ and $\hat{\mu}$ is the retrieved one. Figure 2 presents the obtained results (averaged over 1000 simulations). It is worth mentioning that we randomly generated the capacity μ according to the *Random-Node Generator*⁵ proposed by Havens and Pinar (2017). With respect to $u(\mathbf{x}_j)$, we considered a uniform distribution in the range $[0, 1]$.

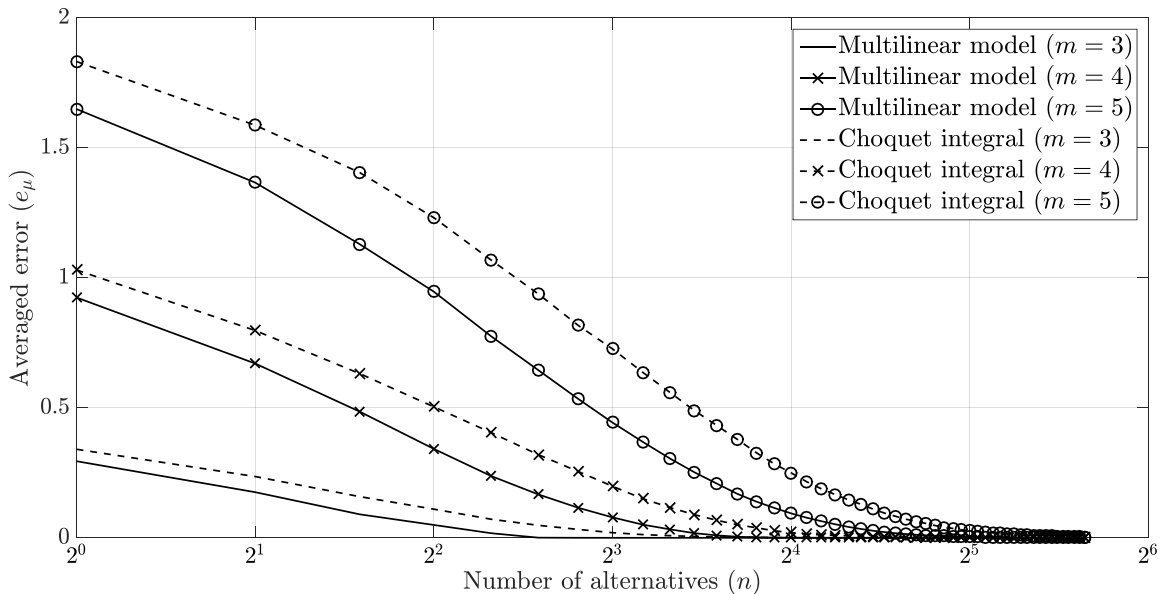


Figure 2: Mean squared error for different number of alternatives.

By taking the multilinear model, one may note that, for 3, 4 and 5 criteria, $e_\mu \approx 0$ when we have

⁴For further details about the quadratic model in the context of Choquet integral, the interested readers may see (Anderson et al., 2014).

⁵Since the analyses conducted by Havens and Pinar (2017) pointed out that the *Random-Node Generator* performs better (e.g. in the resulting distribution of the capacity coefficients) compared to other methods exploited by the authors, we adopted this procedure in our experiments.

at least (2^3) , (2^4) and (2^5) learning data, respectively. However, one needs more samples in order to achieve the same performance by considering the Choquet integral. This can be explained by the fact that, as mentioned in Section 1, only a subset of the capacity coefficients are used when applying the Choquet integral, which depends on the order of the set of evaluations. Therefore, the matrix \mathbf{P} applied in the optimization model may have zero columns, which will not lead to the identification of the associated parameters. In that respect, in order to achieve a similar performance with respect to the multilinear model, one must guarantee that the considered learning data “activates” all the parameters that should be retrieved.

For both aggregation functions, if one does not have enough learning data, the problem is ill-posed, and the retrieved capacity is not necessarily the desired one. In this case, one may consider more information in the optimization model in order to retrieve the correct capacity or, at least, to obtain a set of parameters with a high level of generalization. We address this issue in the next section.

4.2. A supervised approach with regularization

The quadratic problem (21) may be ill-posed, which brings a difficulty in identifying the capacity. In that respect one may use a regularization term in order to deal with this ill-posed optimization problem. For instance, as discussed in the context of Choquet integral (Adeyeba et al., 2015), one may apply the ℓ_1 -norm of μ , expressed by⁶

$$\|\mu\|_1 = \sum_{A \subseteq C} \mu(A). \quad (23)$$

By using this regularization term in our problem, one achieves the following optimization model:

$$\begin{aligned} \min_{\mu} \quad & \frac{1}{2} \mu^T \mathbf{Q} \mu + \mathbf{v}^T \mu + \lambda \|\mu\|_1 \\ \text{s.t.} \quad & \mathbf{L} \mu = [0, 1]^T \\ & \mathbf{M} \mu \leq \mathbf{0} \end{aligned} \quad (24)$$

where λ is a constant. Therefore, the solution of (24) provides a vector μ whose most part of the elements are close to zero. However, as pointed out in (Oliveira et al., 2017), the ℓ_1 -norm regularization is more meaningful when applied to the interaction index $I^S(A)$.

⁶The ℓ_1 -norm of a p -dimensional vector $\alpha \in \mathbb{R}^p$ is given by $\|\alpha\|_1 = \sum_{i=1}^p |\alpha_i|$. Therefore, in Equation (23), since $\mu(A) \geq 0, \forall A \subseteq C$, one may eliminate the absolute value.

In the context of the multilinear model, in order to adjust the aforementioned optimization model by taking into account the set of interaction indices $\mathbf{I}^{\mathcal{B}} = [I^{\mathcal{B}}(\emptyset), I^{\mathcal{B}}(\{1\}), \dots, I^{\mathcal{B}}(\{1, \dots\}), \dots, I^{\mathcal{B}}(C)]^T$ as variables, one replaces the capacity by using (6). Therefore, one obtains the following optimization problem⁷:

$$\begin{aligned} \min_{\mathbf{I}^{\mathcal{B}}} \quad & \frac{1}{2} (\mathbf{I}^{\mathcal{B}})^T \mathbf{Q}' \mathbf{I}^{\mathcal{B}} + (\mathbf{v}')^T \mathbf{I}^{\mathcal{B}} + \lambda \|\mathbf{I}^{\mathcal{B}}\|_1 \\ \text{s.t.} \quad & \mathbf{L}' \mathbf{I}^{\mathcal{B}} = [0, 1]^T \\ & \mathbf{M}' \mathbf{I}^{\mathcal{B}} \leq \mathbf{0} \end{aligned} \tag{25}$$

where $\mathbf{Q}' = \mathbf{S}^T \mathbf{Q} \mathbf{S}$, $\mathbf{v}' = \mathbf{S}^T \mathbf{v}$, $\mathbf{L}' = \mathbf{L} \mathbf{S}$, $\mathbf{M}' = \mathbf{M} \mathbf{S}$ and \mathbf{S} is the transformation matrix from $\mathbf{I}^{\mathcal{B}}$ to μ (i.e. $\mu = \mathbf{S} \mathbf{I}^{\mathcal{B}}$), given by

$$\mathbf{S} = \begin{bmatrix} \left(\frac{1}{2}\right)^{|\emptyset|} (-1)^{|\emptyset \setminus \emptyset|} & \left(\frac{1}{2}\right)^{|\{1\}|} (-1)^{|\{1\} \setminus \emptyset|} & \dots & \left(\frac{1}{2}\right)^{|C|} (-1)^{|C \setminus \emptyset|} \\ \left(\frac{1}{2}\right)^{|\emptyset|} (-1)^{|\emptyset \setminus \{1\}|} & \left(\frac{1}{2}\right)^{|\{1\}|} (-1)^{|\{1\} \setminus \{1\}|} & \dots & \left(\frac{1}{2}\right)^{|C|} (-1)^{|C \setminus \{1\}|} \\ \vdots & \vdots & \ddots & \vdots \\ \left(\frac{1}{2}\right)^{|\emptyset|} (-1)^{|\emptyset \setminus C|} & \left(\frac{1}{2}\right)^{|\{1\}|} (-1)^{|\{1\} \setminus C|} & \dots & \left(\frac{1}{2}\right)^{|C|} (-1)^{|C \setminus C|} \end{bmatrix}.$$

One may note that (25) is not quadratic anymore. However, by considering a set of auxiliary variables (Vanderbei, 2014), it is possible to turn (25) into a quadratic problem. Moreover, we can include a vector $\mathbf{z}_\gamma = [\mathbf{z}_\gamma(\emptyset), \mathbf{z}_\gamma(\{1\}), \dots, \mathbf{z}_\gamma(\{1, \dots\}), \dots, \mathbf{z}_\gamma(C)]^T$, such that

$$\mathbf{z}_\gamma(A) = \begin{cases} 1, & \text{if } |A| > \gamma \\ 0, & \text{otherwise} \end{cases},$$

which controls the application of the ℓ_1 -norm in a subset of all the interaction indices. Indeed, consider that $\mathbf{I}^{\mathcal{B}} = \mathbf{I}^{\mathcal{B}^+} - \mathbf{I}^{\mathcal{B}^-}$, with both $\mathbf{I}^{\mathcal{B}^+}, \mathbf{I}^{\mathcal{B}^-} \geq \mathbf{0}$, and $|\mathbf{I}^{\mathcal{B}}| = \mathbf{I}^{\mathcal{B}^+} + \mathbf{I}^{\mathcal{B}^-}$. Moreover, define

$$\tilde{\mathbf{I}}^{\mathcal{B}} = \begin{bmatrix} \mathbf{I}^{\mathcal{B}^+} \\ \mathbf{I}^{\mathcal{B}^-} \end{bmatrix}, \quad \mathbf{Q}'' = \begin{bmatrix} \mathbf{Q}' & -\mathbf{Q}' \\ -\mathbf{Q}' & \mathbf{Q}' \end{bmatrix}, \quad \mathbf{v}'' = \begin{bmatrix} \lambda \mathbf{z}_\gamma + \mathbf{v}' \\ \lambda \mathbf{z}_\gamma - \mathbf{v}' \end{bmatrix}, \quad \mathbf{L}'' = \begin{bmatrix} \mathbf{L}' & -\mathbf{L}' \end{bmatrix} \quad \text{and} \quad \mathbf{M}'' = \begin{bmatrix} \mathbf{M}' & -\mathbf{M}' \end{bmatrix}.$$

⁷It is worth mentioning that, for $\lambda = 0$ and $I^{\mathcal{B}}(A) = 0$ for all A such that $|A| \geq 3$, (25) leads to the optimization model by means of a 2-additive capacity.

The optimization problem described in (25) may be expressed, in a quadratic fashion, as follows:

$$\begin{aligned}
\min_{\tilde{\mathbf{I}}^{\mathcal{B}}} \quad & \frac{1}{2} (\tilde{\mathbf{I}}^{\mathcal{B}}) \mathbf{Q}'' \tilde{\mathbf{I}}^{\mathcal{B}} + (\mathbf{v}'')^T \tilde{\mathbf{I}}^{\mathcal{B}} \\
\text{s.t.} \quad & \mathbf{L}'' \tilde{\mathbf{I}}^{\mathcal{B}} = [0, 1]^T \\
& \mathbf{M}'' \tilde{\mathbf{I}}^{\mathcal{B}} \leq \mathbf{0} \\
& \tilde{\mathbf{I}}^{\mathcal{B}} \geq \mathbf{0}
\end{aligned} \tag{26}$$

The rationality behind the use of the ℓ_1 -norm regularization in the interaction index $I^{\mathcal{B}}(A)$ relies on the search for a simpler model, in which a considerable number of parameters tends to be equal to zero. Therefore, one achieves a sparse solution for $\mathbf{I}^{\mathcal{B}}$, whose sparsity level depends on the λ value. For example, if we consider $\gamma = 1$, the solution of (26) promotes a set of interaction indices in which $I^{\mathcal{B}}(A)$, $|A| \geq 2$, is close to zero. Conversely, if we consider $\gamma = 2$, the solution of (26) promotes a set of interaction indices in which $I^{\mathcal{B}}(A)$, $|A| \geq 3$, is close to zero. Therefore, one reduces the flexibility of the model to arbitrarily adjust the set of parameters in the identification problem. The price to be paid is that, by decreasing the level of flexibility, one may increase the representation error E between the obtained evaluations and the desired ones (expressed in Equation (20)). We provide a further discussion on this topic in the experiment on real data.

Although a simpler model may achieve a larger value of E in the training step, depending on the addressed MCDM problem, the retrieved capacity may be close to the correct one. Moreover, one may also find parameters that lead to a better generalization by applying the retrieved capacity in a dataset other than the one used in the training step. These insights are further discussed in the next section.

5. Numerical experiments

5.1. Experiments with synthetic data

The first experiment comprises the application of the considered methods with a set of synthetic data. For instance, consider the following notation:

- WRE: Supervised approach without regularization, expressed in Equation (21);
- RE2: Supervised approach with regularization, expressed in Equation (26), with $\gamma = 2$;
- 2AD: Supervised approach by means of a 2-additive capacity, expressed in Equation (25), with $\lambda = 0$ and $I^{\mathcal{B}}(A) = 0$ for all A such that $|A| \geq 3$;

- RE1: Supervised approach with regularization, expressed in Equation (26), with $\gamma = 1$;
- WAM: Weighted arithmetic mean (additive function).

The data were generated according to the procedure described in (Grabisch, 1995). We considered $n = 81$ learning data, with $m = 4$ criteria. For each evaluation $u_{i,j}$, we randomly selected a value belonging to the set $\{0, 0.5, 1\}$. Therefore, the global evaluation of alternative \mathbf{x}_j was obtained by the following expression:

$$y(u(\mathbf{x}_j)) = F_{ML}(u(\mathbf{x}_j)) + g, \quad (27)$$

where g represents an additive Gaussian noise with variance $\sigma^2 = 0.0125$.

Let us consider the weights $\lambda_{RE1} = 1$ and $\lambda_{RE2} = 0.015$ for RE1 and RE2 methods, respectively. It is worth mentioning that, in all analysis conducted in this paper, the adopted λ_{RE1} and λ_{RE2} were experimentally defined. The application of WRE, RE2, 2AD, RE1 and WAM methods in the capacity identification problem leads to the representation errors $E_{WRE} = 0.0112$, $E_{RE2} = 0.0116$, $E_{2AD} = 0.0118$, $E_{RE1} = 0.0616$ and $E_{WAM} = 0.2254$, respectively. The retrieved capacities and interaction indices are described in Table 1, which also presents the mean squared error between the correct capacity and the retrieved ones, given by

$$\epsilon_\mu = \frac{1}{2^m - 2} \sum_{A \subseteq C, A \neq \emptyset, A \neq C} (\mu(A) - \hat{\mu}(A))^2, \quad (28)$$

and the mean squared error between the correct interaction indices and the retrieved ones, given by

$$\epsilon_{I^B} = \frac{1}{2^m} \sum_{A \in C} \left(I^B(A) - \hat{I}^B(A) \right)^2. \quad (29)$$

With respect to the WAM method, one obtains the set of weights $\mathbf{w} = [0.0825, 0.1850, 0.1581, 0.5744]$.

As expected, the WRE method achieved the lower representation error E . However, if we compare the considered approaches under the light of the retrieved capacity, both RE2 and 2AD methods provided the better results (lower values of both ϵ_μ and ϵ_{I^B}).

In order to further exploit this analysis, let us verify the generalization ability of the considered approaches. For instance, we divided the samples in two sets. The first set is used in the training step, which leads to the parameters identification, and the second one is used in the test, which validates the retrieved capacity. Assume that one has 12 samples for training (a subset of all samples less than $2^m - 2$ criteria) and that the other 69 samples will be used for test. Moreover,

Table 1: Retrieved capacity and interaction indices (all synthetic learning data).

Capacity $\mu(A)$						Interaction index $I^B(A)$					
A	Correct	WRE	RE2	2AD	RE1	A	Correct	WRE	RE2	2AD	RE1
$\{\emptyset\}$	0	0	0	0	0	$\{\emptyset\}$	0.5511	0.5518	0.5519	0.5518	0.5263
$\{1\}$	0.1	0.1029	0.0991	0.0994	0.0904	$\{1\}$	0.0743	0.0750	0.0770	0.0776	0.0802
$\{2\}$	0.2105	0.2094	0.2066	0.2050	0.1988	$\{2\}$	0.1610	0.1581	0.1595	0.1610	0.1722
$\{3\}$	0.2353	0.2441	0.2334	0.2298	0.1843	$\{3\}$	0.1811	0.1792	0.1792	0.1797	0.1685
$\{4\}$	0.6667	0.6823	0.6749	0.6728	0.6316	$\{4\}$	0.5796	0.5786	0.5810	0.5818	0.5791
$\{1, 2\}$	0.3	0.2992	0.2994	0.2988	0.2891	$\{1, 2\}$	-0.0082	0.0007	-0.0056	-0.0056	≈ 0
$\{1, 3\}$	0.3235	0.3212	0.3241	0.3222	0.2746	$\{1, 3\}$	-0.0093	-0.0099	-0.0083	-0.0070	≈ 0
$\{1, 4\}$	0.7333	0.7328	0.7438	0.7412	0.7016	$\{1, 4\}$	-0.0297	-0.0247	-0.0295	-0.0311	-0.0204
$\{2, 3\}$	0.4211	0.4216	0.4219	0.4225	0.3830	$\{2, 3\}$	-0.0201	-0.0145	-0.0123	-0.0123	≈ 0
$\{2, 4\}$	0.8070	0.7971	0.8053	0.8077	0.7773	$\{2, 4\}$	-0.0644	-0.0654	-0.0698	-0.0702	-0.0531
$\{3, 4\}$	0.8235	0.8152	0.8205	0.8217	0.7844	$\{3, 4\}$	-0.0725	-0.0798	-0.0820	-0.0809	-0.0315
$\{1, 2, 3\}$	0.5	0.5010	0.5063	0.5093	0.4734	$\{1, 2, 3\}$	0.0009	-0.0115	≈ 0	0	≈ 0
$\{1, 2, 4\}$	0.8667	0.8735	0.8692	0.8704	0.8473	$\{1, 2, 4\}$	0.0033	0.0121	0.0014	0	≈ 0
$\{1, 3, 4\}$	0.8824	0.8833	0.8810	0.8831	0.8544	$\{1, 3, 4\}$	0.0038	0.0165	≈ 0	0	≈ 0
$\{2, 3, 4\}$	0.9474	0.9445	0.9444	0.9442	0.9300	$\{2, 3, 4\}$	0.0080	0.0195	0.0116	0	≈ 0
$\{1, 2, 3, 4\}$	1	1	1	1	1	$\{1, 2, 3, 4\}$	-0.0006	-0.0537	≈ 0	0	≈ 0
Error ϵ_μ ($\times 10^{-4}$)	0.4002	0.1911	0.2005	0.2005	9.7187	Error ϵ_{I^B} ($\times 10^{-4}$)	2.2202	0.1480	0.1805	2.1704	

let us define $\lambda_{RE1} = 0.1$ and $\lambda_{RE2} = 0.0025$. Figures 3a and 3b present boxplots⁸ of the obtained representation error E (divided by the number of samples and over 10001 simulations) in training and test steps, respectively. One may remark that, although the WRE method provides the lower E value in training step, the validation of the obtained capacity in the test does not lead to the better results. In this case, the method that achieved the higher levels of generalization were the RE2 and the 2AD.

In Table 2, we show the retrieved capacity and interaction indices (based on the median representation error E in training step over the 10001 simulations). We also present the errors with respect to the estimation of the capacity and interaction indices. The application of the WAM leads to $\mathbf{w} = [0.0997, 0.1860, 0.1870, 0.5273]$. Similarly as obtained by considering all the learning data in the training step, the application of the WRE does not lead to the lower values of ϵ_μ and ϵ_{I^B} .

⁸In each box, the central mark, the top edge and the bottom edge represent the median (q_2), the 75th percentile (q_3) and the 25th percentile (q_1), respectively. Moreover, the whiskers extend until the extremes points that lies in the range $[q_3 - 1.5(q_3 - q_1), q_1 + 1.5(q_3 - q_1)]$, which covers 99,3% of the points, approximately, if the data are normally distributed. The points outside this range are considered as outliers and were removed from the boxplot.

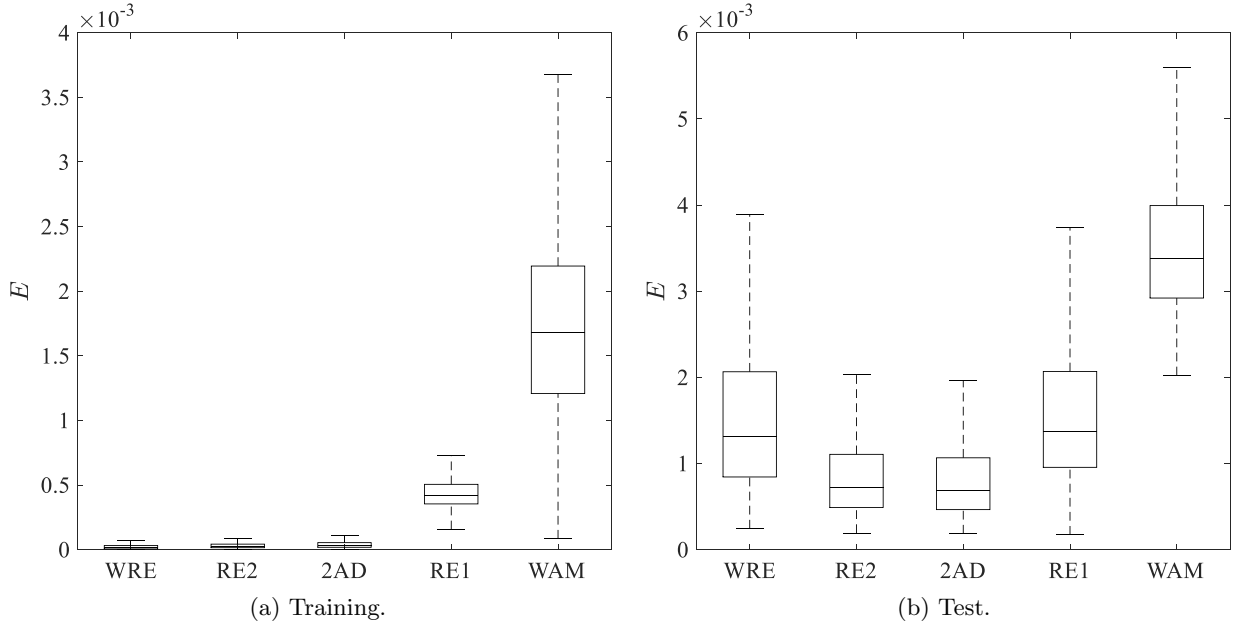


Figure 3: Representation error E based on a subset of the synthetic learning data.

5.2. Experiments with real data

An important issue in ergonomics is the mental workload (Young et al., 2015), which can be measured by using the well-known NASA Task Load Index (NASA-TLX) (Hart and Staveland, 1988). In the application of this procedure, after performing a task, the user provides subjective evaluations (in the range $[0, 100]$) based on six sources: mental demand, physical demand, temporal demand, performance, effort and frustration. Then, the user also provides 15 pairwise comparisons among the six sources, which indicate that a specific source contributes to the workload more than another one. Finally, based on these pairwise comparisons, one determines the “importance” of each source in the mental workload (the weight associated with the source) and obtains the global evaluation by means of a weighted arithmetic mean.

In view of the limitations of the WAM, some works investigated the application of a capacity (Raufaste et al., 2001; Grabisch et al., 2006) to provide the global mental workload evaluation. For instance, other than the six subjective evaluations, the set of users also provide a subjective global evaluation of the task (also in the range $[0, 100]$), which is used as a learning data to perform capacity identification. Therefore, a comparison between the WAM and the Choquet integral can be exploited in order to verify which aggregation function can better represent the information provided by the users.

In this paper, we used the dataset described in (Raufaste et al., 2001), which comprises the

Table 2: Retrieved capacity and interaction indices (subset of the synthetic learning data).

Capacity $\mu(A)$						Interaction index $I^B(A)$					
A	Correct	WRE	RE2	2AD	RE1	A	Correct	WRE	RE2	2AD	RE1
$\{\emptyset\}$	0	0	0	0	0	$\{\emptyset\}$	0.5511	0.5508	0.5523	0.5477	0.5082
$\{1\}$	0.1	0.0526	0.0926	0.0971	0.0984	$\{1\}$	0.0743	0.0842	0.0857	0.0497	0.0984
$\{2\}$	0.2105	0.2038	0.2839	0.1969	0.1741	$\{2\}$	0.1610	0.1507	0.1983	0.1570	0.1577
$\{3\}$	0.2353	0.3259	0.1465	0.1987	0.1092	$\{3\}$	0.1811	0.2021	0.1314	0.2194	0.1091
$\{4\}$	0.6667	0.6597	0.6879	0.6982	0.6511	$\{4\}$	0.5796	0.5504	0.5836	0.5738	0.6348
$\{1, 2\}$	0.3	0.3859	0.3747	0.2941	0.2725	$\{1, 2\}$	-0.0082	0.0896	-0.0009	≈ 0	≈ 0
$\{1, 3\}$	0.3235	0.3872	0.3178	0.2981	0.2076	$\{1, 3\}$	-0.0093	-0.0553	0.0797	0.0023	≈ 0
$\{1, 4\}$	0.7333	0.6755	0.6879	0.6982	0.7496	$\{1, 4\}$	-0.0297	-0.0231	-0.0907	-0.0971	≈ 0
$\{2, 3\}$	0.4211	0.3811	0.3492	0.4510	0.2831	$\{2, 3\}$	-0.0201	-0.1097	-0.0812	0.0555	-0.0002
$\{2, 4\}$	0.8070	0.7115	0.8827	0.7598	0.7926	$\{2, 4\}$	-0.0644	-0.0352	-0.0882	-0.1353	-0.0326
$\{3, 4\}$	0.8235	0.8373	0.8057	0.8806	0.7603	$\{3, 4\}$	-0.0725	-0.0556	-0.0277	-0.0163	≈ 0
$\{1, 2, 3\}$	0.5	0.4679	0.5188	0.5505	0.3815	$\{1, 2, 3\}$	0.0009	-0.1316	≈ 0	0	≈ 0
$\{1, 2, 4\}$	0.8667	0.9085	0.8827	0.7598	0.8910	$\{1, 2, 4\}$	0.0033	0.0242	0.0018	0	≈ 0
$\{1, 3, 4\}$	0.8824	0.8652	0.8863	0.8829	0.8587	$\{1, 3, 4\}$	0.0038	-0.0242	0.0019	0	≈ 0
$\{2, 3, 4\}$	0.9474	0.9500	0.9193	0.9977	0.9016	$\{2, 3, 4\}$	0.0080	0.1819	0.0001	0	≈ 0
$\{1, 2, 3, 4\}$	1	1	1	1	1	$\{1, 2, 3, 4\}$	-0.0006	-0.0551	≈ 0	0	≈ 0
Error ϵ_μ ($\times 10^{-2}$)	0.2809	0.2435	0.1951	0.5177		Error ϵ_{IB} ($\times 10^{-2}$)	0.4650	0.1380	0.1304	0.1154	

subjective evaluations (over the six sources and the global one) provided by a set of 143 users⁹. Similarly to the experiments with synthetic data, we firstly apply the considered approaches by taking into account all the 143 samples as the learning data. Before setting the regularization weights used in RE1 and RE2 methods, let us investigate the performance of the considered approaches for different values of λ . Figure 4 illustrates the obtained representation errors. Clearly, in the case of WRE, 2AD and WAM, E is not affected by λ . However, as highlighted in Section 4.2, there is a trade-off between regularization and representation error when RE1 and RE2 methods are applied. In both cases, if one sets λ close to zero, which practically eliminates the regularization term, the methods achieve the same performance as the WRE. However, if one adopts a large value of λ , which means that one considers that the minimization of the ℓ_1 -norm is very important in the optimization model, the parameters associated with this regularization term will be approximately

⁹Originally, there were 188 samples, however, one eliminates 45 due to inconsistencies.

zero. Therefore, RE1 and RE2 will converge to the WAM and 2AD methods, respectively.

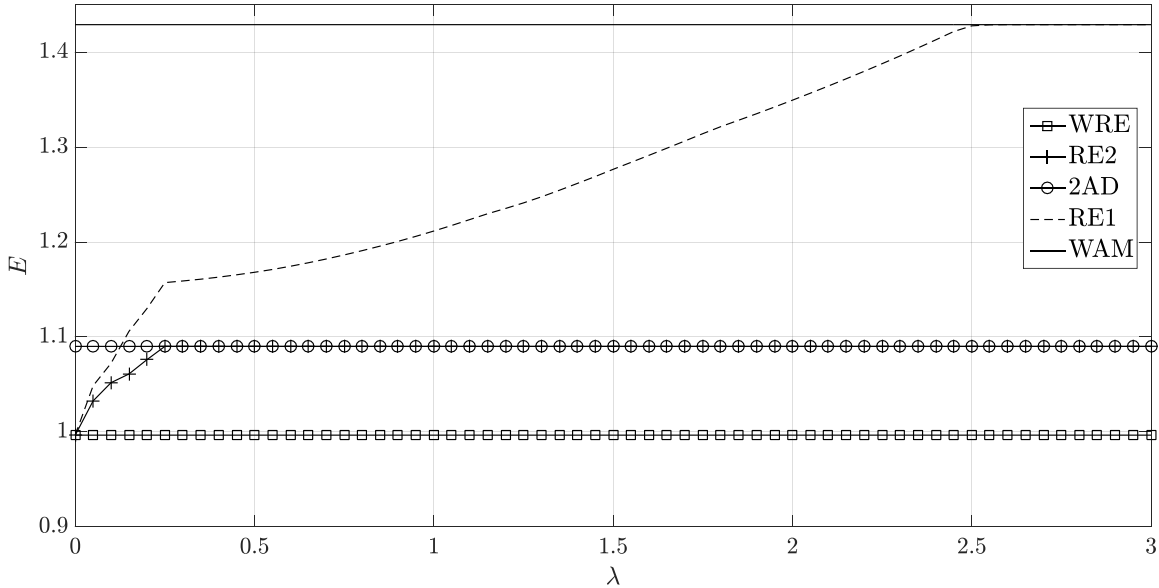


Figure 4: Obtained representation error E for different values of λ .

Assume $\lambda_{RE1} = 1$ and $\lambda_{RE2} = 0.1$ for RE1 and RE2 methods, respectively. The application of WRE, RE2, 2AD, RE1 and WAM methods in the capacity identification problem leads to the representation errors $E_{WRE} = 0.9966$, $E_{RE2} = 1.0519$, $E_{2AD} = 1.0901$, $E_{RE1} = 1.2116$ and $E_{WAM} = 1.4290$, respectively. Therefore, one may note that the methods that take into account interactions among criteria achieved values of representation error E that are considerably lower compared to the application of the WAM. This means that the aggregation conducted by WAM may not be sufficient to satisfy the information provided by the users.

The retrieved capacities and interaction indices are described in Table 3, as well as the associated error. Since we have $m = 6$ criteria and, therefore, $2^6 = 64$ parameters, for the purpose of illustrating the obtained results, we only show the ones such that $1 \leq |A| \leq 2$. It is worth mentioning that the application of the WAM leads to $\mathbf{w} = [0.5056, 0, 0.0622, 0.3056, 0.0854, 0.0412]$.

Aiming at verifying the generalization ability of the considered approaches, let us consider a training step with a subset of all samples. For instance, consider that one has 100 samples for training and that the other 43 samples will be used for test. Moreover, let us define $\lambda_{RE1} = 0.1$ and $\lambda_{RE2} = 0.05$. Figures 5a and 5b present the boxplots of the obtained representation error E (divided by the number of samples and over 1001 simulations) in training and test steps, respectively. One may also remark here that, although the WRE method provides the lower E value in training step, the validation of the obtained capacity in the test step does not lead to the better results. One also

Table 3: Retrieved capacity and interaction indices (all real learning data).

Capacity $\mu(A)$					Interaction index $I^B(A)$				
A	WRE	RE2	2AD	RE1	A	WRE	RE2	2AD	RE1
{1}	0.5580	0.5194	0.4953	0.4671	{1}	0.4414	0.4256	0.4140	0.4608
{2}	0.0001	0.1423	0.1821	0.0375	{2}	0.0979	0.0982	0.0995	0.0352
{3}	0.1344	0.1440	0.1310	0.1076	{3}	0.1194	0.0726	0.0740	0.0648
{4}	≈ 0	0.3032	0.4077	0.3179	{4}	0.2933	0.2977	0.2741	0.3132
{5}	0.0001	0.0044	0.0054	0.0824	{5}	0.1074	0.0787	0.0647	0.0674
{6}	0.0001	0.0168	0.0353	0.1171	{6}	0.0906	0.0840	0.0737	0.0586
{1, 2}	0.5584	0.6275	0.5946	0.5046	{1, 2}	-0.0133	-0.0360	-0.0828	≈ 0
{1, 3}	0.6546	0.6536	0.6262	0.5715	{1, 3}	-0.1098	-0.0107	≈ 0	-0.0031
{1, 4}	0.6196	0.6388	0.7113	0.7756	{1, 4}	-0.1879	-0.1989	-0.1917	-0.0094
{1, 5}	0.5581	0.5238	0.5007	0.5495	{1, 5}	-0.0953	-0.0151	≈ 0	≈ 0
{1, 6}	0.6547	0.5943	0.6426	0.5842	{1, 6}	0.1175	0.0555	0.1121	≈ 0
{2, 3}	0.3569	0.2864	0.3301	0.1451	{2, 3}	0.0838	0.0001	0.0169	≈ 0
{2, 4}	0.4237	0.4456	0.4992	0.3555	{2, 4}	-0.0514	-0.0523	-0.0906	≈ 0
{2, 5}	0.0012	0.1467	0.1875	0.1200	{2, 5}	0.0148	≈ 0	≈ 0	≈ 0
{2, 6}	0.0011	0.1591	0.2087	0.1501	{2, 6}	-0.0432	-0.0540	-0.0087	-0.0045
{3, 4}	0.4563	0.3279	0.4297	0.4255	{3, 4}	-0.1709	-0.1196	-0.1089	≈ 0
{3, 5}	0.1345	0.1483	0.1360	0.1900	{3, 5}	-0.0117	-0.0003	-0.0004	≈ 0
{3, 6}	0.1344	0.1483	0.1447	0.1422	{3, 6}	-0.0397	-0.0134	-0.0216	-0.0824
{4, 5}	0.6464	0.5193	0.5371	0.4004	{4, 5}	0.1745	0.1530	0.1240	≈ 0
{4, 6}	0.4563	0.5115	0.4430	0.4351	{4, 6}	0.0769	0.0958	≈ 0	≈ 0
{5, 6}	0.0001	0.0169	0.0357	0.1694	{5, 6}	-0.0643	-0.0477	-0.0050	-0.0302

obtained higher generalization levels with the application of RE2 and 2AD methods.

In Table 4, we show the retrieved capacity and interaction indices (based on the median representation error E in training step over the 1001 simulations). With the application of WAM, one obtains $\mathbf{w} = [0.5382, 0.0287, 0.0799, 0.2476, 0.0537, 0.0519]$.

6. Conclusions

Aggregation functions that take into account interaction among criteria, such as the Choquet integral and the multilinear model, can be useful in several practical situations. Differently from the Choquet integral, few works in the literature exploit theoretical aspects or applications of the

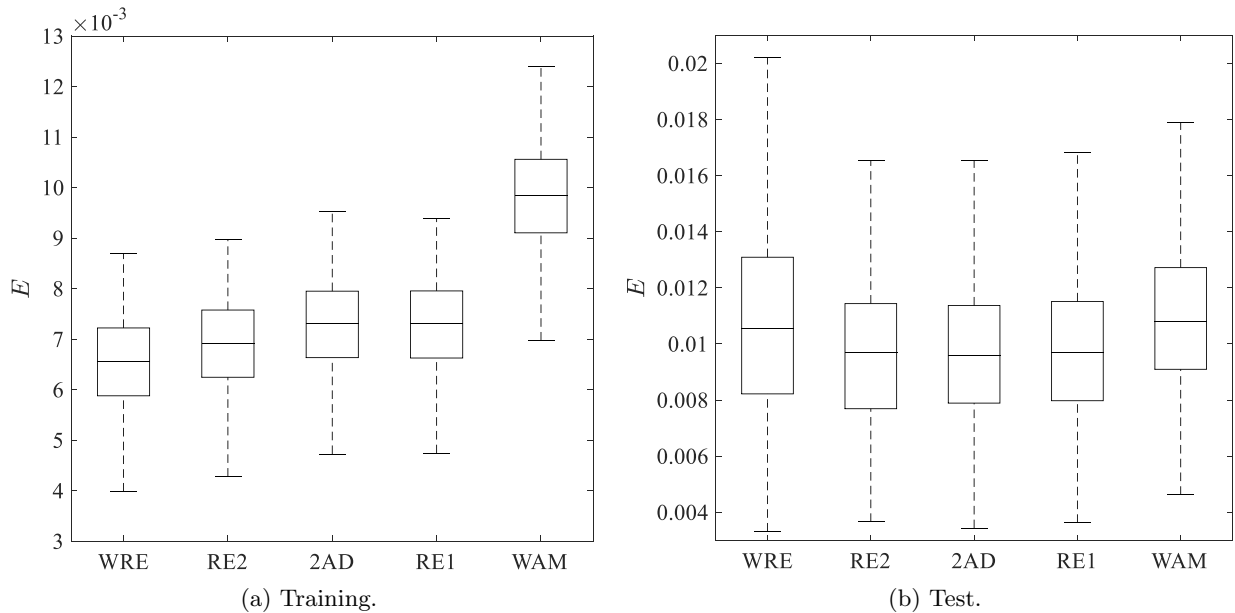


Figure 5: Representation error E based on a subset of the real learning data.

multilinear model. Therefore, this paper proposed to address some issues in MCDM problems by means of the multilinear model and open the path for future researches on this subject.

As a first analysis, we exploited the concept of a 2-additive capacity in the multilinear model and provide an analytical expression. Moreover, it was possible to remark some similarities between the 2-additive Choquet integral and the 2-additive multilinear model. In both functionals, one applies an additive, a disjunctive and a conjunctive term to aggregate the set of evaluations. Therefore, we could generalize them into a single expression.

We also addressed the problem of capacity identification in a supervised fashion. As a first remark, we noted that, in order that the Choquet integral achieves the same performance of the multilinear model, one needs to ensure that the considered learning data will lead to the identification of all coefficients of the capacity. Otherwise, some parameters may not be retrieved by the optimization model. Moreover, in the absence of enough data, the problem is ill-posed, and one should consider more information to perform capacity identification. In that respect, we exploited a supervised approach with regularization, which can lead to a simpler model compared to the one obtained without the use of this additional term. In this case, there is a loss of performance with respect to the representation error in the optimization model but one can retrieve a capacity close to the correct one or, at least, with a higher level of generalization when applied in a new dataset.

As mentioned in Section 5.1, we experimentally defined the values of regularization weights.

Table 4: Retrieved capacity and interaction indices (subset of the real learning data).

Capacity $\mu(A)$					Interaction index $I^B(A)$				
A	WRE	RE2	2AD	RE1	A	WRE	RE2	2AD	RE1
{1}	0.0350	0.5282	0.4917	0.3968	{1}	0.5197	0.3470	0.4832	0.4459
{2}	0.0313	0.1544	0.2262	0.1503	{2}	0.1023	0.1277	0.1131	0.1222
{3}	0.1384	0.0286	0.2025	0.1737	{3}	0.1262	0.0750	0.1013	0.0868
{4}	≈ 0	0.4506	0.3009	0.3808	{4}	0.2488	0.3433	0.1987	0.2535
{5}	0.0313	≈ 0	≈ 0	0.0374	{5}	0.0549	0.0813	0.0266	0.0537
{6}	0.0313	0.0268	≈ 0	0.0374	{6}	0.1166	0.0761	0.0770	0.0791
{1, 2}	0.5565	0.5365	0.7177	0.6295	{1, 2}	-0.0009	-0.1463	-0.0001	≈ 0
{1, 3}	0.6714	0.5552	0.5665	0.4953	{1, 3}	-0.0944	-0.0017	-0.1276	-0.0751
{1, 4}	0.7383	0.6400	0.7926	0.6778	{1, 4}	0.0065	-0.3390	≈ 0	-0.0998
{1, 5}	0.5565	0.5284	0.4917	0.5041	{1, 5}	0.0589	≈ 0	≈ 0	0.0699
{1, 6}	0.6714	0.6794	0.6024	0.6374	{1, 6}	0.1475	0.1242	0.1108	0.1209
{2, 3}	0.4463	0.3569	0.4287	0.3240	{2, 3}	0.0305	0.0986	0.0001	≈ 0
{2, 4}	0.3780	0.5972	0.3010	0.4748	{2, 4}	-0.1510	-0.0842	-0.2261	-0.0564
{2, 5}	0.0313	0.1572	0.2262	0.1878	{2, 5}	-0.0596	-0.0001	≈ 0	≈ 0
{2, 6}	0.0313	0.1830	0.2262	0.1878	{2, 6}	0.0043	0.0008	≈ 0	-0.0824
{3, 4}	0.4463	0.5248	0.4285	0.4560	{3, 4}	-0.1642	-0.0520	-0.0748	-0.0986
{3, 5}	0.1384	0.0287	0.2025	0.2111	{3, 5}	-0.0593	-0.0233	≈ 0	≈ 0
{3, 6}	0.1384	0.0286	0.2025	0.2111	{3, 6}	-0.0936	-0.0268	≈ 0	≈ 0
{4, 5}	0.3753	0.6363	0.3542	0.4183	{4, 5}	0.0386	0.1610	0.0533	≈ 0
{4, 6}	0.4463	0.4778	0.3442	0.4183	{4, 6}	-0.0247	≈ 0	0.0433	≈ 0
{5, 6}	0.0313	0.0269	≈ 0	0.0374	{5, 6}	-0.0995	-0.0010	≈ 0	-0.0374

Therefore, as a future work, one aims at investigating an automatic procedure to set these parameters. A remark on this issue is that, since the cost function comprises one part related to the global evaluations and another one associated to the interaction indices, one needs to take into account this difference in terms of the nature of each part. Another future perspective lies on the development of a non-supervised approach, whose goal is to perform the capacity identification based only on the set of criteria evaluations and on an assumption/information about these data, without the use of the overall values in the learning procedure. For instance, as exploited by existing approaches (Duarte, 2018), one may assume that the criteria are correlated and associate a similarity measure between criteria to the interaction indices.

Acknowledgments.

This work was supported by the São Paulo Research Foundation (FAPESP, grant numbers 2016/21571-4 and 2017/23879-9) and the National Council for Scientific and Technological Development (CNPq, grant number 311357/2017-2). Moreover, the authors would like to thank the *Institut National de Recherche et de Sécurité* (INRS) for allowing them to use, for the present work, the data collected in the framework of a previous collaboration (convention N° 5001053).

References

- Adeyeba, T. A., Anderson, D. T., & Havens, T. C. (2015). Insights and characterization of l_1 -norm based sparsity learning of a lexicographically encoded capacity vector for the Choquet integral. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE.
- Anderson, D. T., Price, S. R., & Havens, T. C. (2014). Regularization-based learning of the Choquet integral. In *2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)* (pp. 2519–2526). Beijing, China: IEEE.
- Angilella, S., Corrente, S., & Greco, S. (2015). Stochastic multiobjective acceptability analysis for the Choquet integral preference model and the scale construction problem. *European Journal of Operational Research*, *240*, 172–182.
- Angilella, S., Greco, S., & Matarazzo, B. (2010). Non-additive robust ordinal regression: A multiple criteria decision model based on the Choquet integral. *European Journal of Operational Research*, *201*, 277–288.
- Banzhaf, J. F. (1965). Weighted voting doesn't work: A mathematical analysis. *Rutgers Law Review*, *19*, 317–343.
- Beliakov, G., Pradera, A., & Calvo, T. (2007). *Aggregation functions: A guide for practitioners*. Germany: Springer.
- Blackorby, C., Primont, D., & Russell, R. R. (1978). *Duality, separability, and functional structure: Theory and economic applications*. North-Holland.
- Bottero, M., Ferretti, V., Figueira, J. R., Greco, S., & Roy, B. (2018). On the Choquet multiple criteria preference aggregation model: Theoretical and practical insights from a real-world application. *European Journal of Operational Research*, *271*, 120–140.

- Bouyssou, D., & Pirlot, M. (2004). Preferences for multi-attributed alternatives: Traces, dominance, and numerical representations. *Journal of Mathematical Psychology*, *48*, 167–185.
- Choquet, G. (1954). Theory of capacities. *Annales de l'Institut Fourier*, *5*, 131–295.
- Combarro, E. F., & Miranda, P. (2006). Identification of fuzzy measures from sample data with genetic algorithms. *Computers & Operations Research*, *33*, 3046–3066.
- Duarte, L. T. (2018). A novel multicriteria decision aiding method based on unsupervised aggregation via the Choquet integral. *IEEE Transactions on Engineering Management*, *65*, 293–302.
- Feyzioglu, O., & Büyüközkan, G. (2010). Evaluation of green suppliers considering decision criteria dependencies. In M. Ehrgott, B. Naujoks, T. Stewart, & J. Wallenius (Eds.), *Multiple Criteria Decision Making for Sustainable Energy and Transportation Systems. Lecture Notes in Economics and Mathematical Systems* (pp. 145–154). Berlin, Heidelberg: Springer.
- Figueira, J., Greco, S., & Ehrgott, M. (Eds.) (2016). *Multiple criteria decision analysis: State of the art surveys*. (2nd ed.). New York: Springer's International Series in Operations Research & Management Science.
- Grabisch, M. (1995). A new algorithm for identifying fuzzy measures and its application to pattern recognition. In *International Joint Conference of the Fourth IEEE International Conference on Fuzzy Systems and The Second International Fuzzy Engineering Symposium* (pp. 145–150). Yokohama, Japan.
- Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European Journal of Operational Research*, *89*, 445–456.
- Grabisch, M. (1997). k -order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, *92*, 167–189.
- Grabisch, M. (2000). A graphical interpretation of the Choquet integral. *IEEE Transactions on Fuzzy Systems*, *8*, 627–631.
- Grabisch, M. (2016). *Set Functions, games and capacities in decision making*. Switzerland: Springer International Publishing.

- Grabisch, M., Kojadinovic, I., & Meyer, P. (2008). 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*, 766–785.
- Grabisch, M., & Labreuche, C. (2016). Fuzzy measures and integrals in MCDA. In S. Greco, M. Ehrgott, & J. Figueira (Eds.), *Multiple Criteria Decision Analysis. International Series in Operations Research & Management Science* (pp. 553–603). New York, NY.
- Grabisch, M., Marichal, J.-L., & Roubens, M. (2000). Equivalent representations of set functions. *Mathematics of Operations Research*, *25*, 157–178.
- Grabisch, M., Prade, H., Raufaste, E., & Terrier, P. (2006). Application of the Choquet integral to subjective mental workload evaluation. *IFAC Proceedings Volumes*, *39*, 135–140.
- Greco, S., Matarazzo, B., & Słowiński, R. (2004). Axiomatic characterization of a general utility function and its particular cases in terms of conjoint measurement and rough-set decision rules. *European Journal of Operational Research*, *158*, 271–292.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock, & N. Meshkati (Eds.), *Human Mental Workload* (pp. 139–183). North-Holland volume 52 of *Advances in Psychology*.
- Havens, T. C., & Pinar, A. J. (2017). Generating random fuzzy (capacity) measures for data fusion simulations. In *2017 IEEE Symposium Series on Computational Intelligence (IEEE SSCI 2017)* (pp. 1–8). IEEE.
- Keeney, R. L., & Raiffa, H. (1976). *Decisions with multiple objectives*. New York: Wiley.
- Klement, E. P., Mesiar, R., & Pap, E. (2000). *Triangular norms*. Kluwer Academic Publishers.
- Kojadinovic, I. (2004). Estimation of the weights of interacting criteria from the set of profiles by means of information-theoretic functionals. *European Journal of Operational Research*, *155*, 741–751.
- Kojadinovic, I. (2008). Unsupervised aggregation of commensurate correlated attributes by means of the Choquet integral and entropy functionals. *International Journal of Intelligent Systems*, *23*, 128–154.

- Kolesárová, A., Li, J., & Mesiar, R. (2018). k -additive aggregation functions and their characterization. *European Journal of Operational Research*, *265*, 985–992.
- Kolesárová, A., Stupňanová, A., & Beganová, J. (2012). Aggregation-based extensions of fuzzy measures. *Fuzzy Sets and Systems*, *194*, 1–14.
- Li, G., Law, R., Vu, H. Q., & Rong, J. (2013). Discovering the hotel selection preferences of Hong Kong inbound travelers using the Choquet integral. *Tourism Management*, *36*, 321–330.
- Marichal, J.-L., & Roubens, M. (2000). Determination of weights of interacting criteria from a reference set. *European Journal of Operational Research*, *124*, 641–650.
- Miranda, P., Grabisch, M., & Gil, P. (2003). Identification of non-additive measures from sample data. In G. Della Riccia, D. Dubois, R. Kruse, & H. J. Lenz (Eds.), *Planning Based on Decision Theory. International Centre for Mechanical Sciences (Courses and Lectures)* (pp. 43–62). Vienna: Springer.
- Oliveira, H. E., Romano, J. M. T., & Duarte, L. T. (2017). Identificação dos parâmetros da integral de Choquet via uma abordagem baseada em processamento de sinais esparsos (Identification of Choquet integral parameters via an approach based on sparse signal processing). In *XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT 2017)* (pp. 692–696). São Pedro, São Paulo, Brazil.
- Owen, G. (1972). Multilinear extensions of games. *Management Science, Part 2*, *18*, 64–79.
- Pelegrina, G. D., Grabisch, M., Duarte, L. T., & Romano, M. T. (2018). Multilinear model: New issues in capacity identification. In *From Multiple Criteria Decision Aid to Preference Learning (DA2PL'2018) Workshop*. Poznan, Poland. URL: <http://da2pl.cs.put.poznan.pl/programme/detailed-programme/da2pl2018-abstract-10.pdf>.
- Raufaste, E., Terrier, P., & Grabisch, M. (2001). Etude expérimentale de l'applicabilité de modèles d'agrégations flous à l'étude de la charge mentale (Experimental study of the applicability of fuzzy aggregation models to the study of mental workload). In *Journées d'études en Psychologie Ergonomique (EPIQUE'01)* (pp. 171–176). Nantes, France.
- Rota, G.-C. (1964). On the foundations of combinatorial theory I. Theory of Möbius functions. *Wahrscheinlichkeitstheorie verw Gebiete*, *2*, 340–368.

Roubens, M. (1996). Interaction between criteria through the use of fuzzy measures. In *44th Meeting of the European Working Group "Multicriteria Aid for Decisions"*. Brussels, Belgium.

Rowley, H. V., Geschke, A., & Lenzen, M. (2015). A practical approach for estimating weights of interacting criteria from profile sets. *Fuzzy Sets and Systems*, 272, 70–88.

Shapley, L. S. (1953). A value for n-person games. In W. Kuhn, & A. W. Tucker (Eds.), *Annals of mathematics studies: Vol. 28. Contributions to the theory of games, Vol. II* (pp. 307–317). Princeton: Princeton University Press.

Vanderbei, R. J. (2014). *Linear programming: Foundations and extensions*. (4th ed.). New York: Springer Science+Business Media.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, 58, 1–17.

Appendix A. Axioms of a capacity in a 2-additive multilinear model

In terms of the Banzhaf interaction index, the axioms of a capacity may be expressed by

$$\mu(\emptyset) = 0 \rightarrow \mu(\emptyset) = \sum_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus \emptyset|} I^{\mathcal{B}}(B) = \sum_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B|} I^{\mathcal{B}}(B) = \sum_{B \subseteq C} \left(-\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) = 0;$$

$$\mu(C) = 1 \rightarrow \mu(C) = \sum_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus C|} I^{\mathcal{B}}(B) = \sum_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|0|} I^{\mathcal{B}}(B) = \sum_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) = 1;$$

$$\begin{aligned} \mu(\{A \cup i\}) - \mu(A) \geq 0 &\rightarrow \mu(\{A \cup i\}) - \mu(A) = \sum_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus \{A \cup i\}|} I^{\mathcal{B}}(B) - \sum_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus A|} I^{\mathcal{B}}(B) \\ &= \sum_{B \subseteq C} \left(\frac{1}{2}\right)^{|B|} \left[(-1)^{|B \setminus \{A \cup i\}|} - (-1)^{|B \setminus A|} \right] I^{\mathcal{B}}(B) \geq 0. \end{aligned}$$

If $i \notin B$, $(-1)^{|B \setminus \{A \cup i\}|} = (-1)^{|B \setminus A|}$ and, therefore $\left[(-1)^{|B \setminus \{A \cup i\}|} - (-1)^{|B \setminus A|} \right] = 0$. However, if $B \ni i$, we have

$$\begin{aligned}
\sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} \left[(-1)^{|B \setminus \{A \cup i\}|} - (-1)^{|B \setminus A|} \right] I^{\mathcal{B}}(B) &= \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} \left[(-1)^{|B \setminus A|} (-1)^{-1} - (-1)^{|B \setminus A|} \right] I^{\mathcal{B}}(B) \\
&= \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus A|} [(-1) - 1] I^{\mathcal{B}}(B) = (-2) \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus A|} I^{\mathcal{B}}(B)
\end{aligned}$$

and, therefore,

$$\begin{aligned}
\mu(\{A \cup i\}) - \mu(A) \geq 0 &\rightarrow (-2) \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus A|} I^{\mathcal{B}}(B) \geq 0 \\
&\rightarrow \sum_{\substack{B \subseteq C, \\ B \ni i}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus A|} I^{\mathcal{B}}(B) \leq 0, \forall i \in C, \forall A \subseteq C \setminus i.
\end{aligned}$$

In the context of a 2-additive model, these axioms lead to the following:

$$\mu(\emptyset) = 0 \rightarrow \sum_{B \subseteq C, |B| \leq 2} \left(-\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) = I^{\mathcal{B}}(\emptyset) - \frac{1}{2} \sum_i \phi_i^{\mathcal{B}} + \frac{1}{4} \sum_{i, i'} I_{i, i'}^{\mathcal{B}} = 0;$$

$$\mu(C) = 1 \rightarrow \sum_{B \subseteq C, |B| \leq 2} \left(\frac{1}{2}\right)^{|B|} I^{\mathcal{B}}(B) = I^{\mathcal{B}}(\emptyset) + \frac{1}{2} \sum_i \phi_i^{\mathcal{B}} + \frac{1}{4} \sum_{i, i'} I_{i, i'}^{\mathcal{B}} = 1;$$

$$\mu(\{A \cup i\}) - \mu(A) \geq 0 \rightarrow \sum_{\substack{B \subseteq C, \\ B \ni i, \\ |B| \leq 2}} \left(\frac{1}{2}\right)^{|B|} (-1)^{|B \setminus A|} I^{\mathcal{B}}(B) = -\frac{1}{2} \phi_i^{\mathcal{B}} + \frac{1}{4} \left(\sum_{\substack{i' \neq i, \\ i' \notin A}} I_{i, i'}^{\mathcal{B}} - \sum_{\substack{i' \neq i, \\ i' \in A}} I_{i, i'}^{\mathcal{B}} \right) \leq 0$$

$$\rightarrow \phi_i^{\mathcal{B}} - \frac{1}{2} \left(\sum_{\substack{i' \neq i, \\ i' \notin A}} I_{i, i'}^{\mathcal{B}} - \sum_{\substack{i' \neq i, \\ i' \in A}} I_{i, i'}^{\mathcal{B}} \right) \geq 0, \forall i \in C, \forall A \subseteq C \setminus i$$

$$\rightarrow \phi_i^{\mathcal{B}} - \frac{1}{2} \sum_{i' \neq i} |I_{i, i'}^{\mathcal{B}}| \geq 0, \forall i \in C.$$