



**HAL**  
open science

## Interpretation of multicriteria decision making models with interacting criteria

Michel Grabisch, Christophe Labreuche

► **To cite this version:**

Michel Grabisch, Christophe Labreuche. Interpretation of multicriteria decision making models with interacting criteria. *New Perspectives in Multiple Criteria Decision Making*, M. Doumpos, J. R. Figueira, S. Greco and C. Zopounidis (eds), inPress. halshs-02381243

**HAL Id: halshs-02381243**

**<https://shs.hal.science/halshs-02381243>**

Submitted on 26 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.

# Interpretation of multicriteria decision making models with interacting criteria

Michel Grabisch and Christophe Labreuche

**Abstract** We consider general MCDA models with discrete attributes. These models are shown to be equivalent to a multichoice game and we put some emphasis on discrete Generalized Independence Models (GAI), especially those which are 2-additive, that is, limited to terms of at most two attributes. The chapter studies the interpretation of these models. For general MCDA models, we study how to define a meaningful importance index, and propose mainly two kinds of importance indices: the signed and the absolute importance indices. For 2-additive GAI models, we study the issue of the decomposition, which is not unique in general. We show that for a monotone 2-additive GAI model, it is always possible to obtain a decomposition where each term is monotone. This has important consequences on the tractability and interpretability of the model.

## 1 Introduction

Traditionally, MCDA is primarily interested in studying (through characterizations) various preference models, and learning them thanks to dedicated elicitation approaches. However, once the model has been obtained, the work is far from being finished. The end-user is highly interested in having insights on the behaviour of the model, and cannot be satisfied with only a black-box model. It is necessary to be able to provide explanations to the user about the reasons behind the decision taken by the model. We are mainly interested in the interpretability of MCDA model in this chapter, where by *interpretability* we mean the ability to provide general in-

---

Michel Grabisch

Université Paris I – Panthéon-Sorbonne, Paris School of Economics, 106-112 Bd de l'Hôpital, 75013 Paris, France, e-mail: michel.grabisch@univ-paris1.fr

Christophe Labreuche

Thales Research & Technology, 1 Avenue Augustin Fresnel, 91767 Palaiseau, France e-mail: christophe.labreuche@thalesgroup.com

formation on the model (e.g., what are the most important criteria), and not the explanation of a specific decision.

The interpretability of a MCDA model is all the more difficult when the MCDA model is rich and captures subtle and complex decision strategies. This is particularly the case when the interaction among criteria is taken into account. Some well-known MCDA models, such as the Choquet integral w.r.t. a ( $k$ -ary) capacity or the Generalized-Additive Independence (GAI) model can represent interaction among criteria, and we will focus on these models in this chapter.

We define in Section 3 a general interpretation of a utility-based MCDA model with discrete attributes, without any restriction on the type of model. More precisely, this interpretation takes the form of importance indices on the decision attributes, and we basically define two kinds of importance indices: the signed importance index and the absolute importance index. The signed importance index computes the average variation induced by an attribute over the model. It is linear in the MCDA model and is similar to the concept of *value* in cooperative game theory. Hence for a GAI model that takes the form of the sum of utilities over small subsets of the attributes, the importance index can be computed on each term separately. On the other hand, the absolute importance index computes the cumulated variation induced by an attribute over the model. It is in general not linear, but becomes linear when the overall utility is a linear combination with nonnegative coefficients of monotone utilities. The question is then whether we have such property for the GAI model.

We address the question of the monotone decomposition of a GAI model in Section 4, focusing on the 2-additive case. More precisely, if we have a nonnegative and monotone 2-additive GAI model, is it possible without loss of generality to assume that each term in the GAI decomposition is itself nonnegative and monotone? If such property is true, this would allow to interpret directly each term in the GAI decomposition, and as we will explain, the complexity of the learning procedure is greatly reduced. The main result of this section shows that indeed, such a decomposition exists (Theorem 5). However, we do not know in general how to obtain a monotone decomposition, and classical decompositions like the one proposed by Braziunas or the ANOVA decomposition do not yield in general a monotone decomposition. The last part of Section 4 is devoted to this question and gives some hints to solve it.

## 2 Background

### 2.1 Multicriteria decision making and conjoint measurement

(see, e.g., Bouyssou and Pirlot (2016) for more details) We consider a multicriteria decision problem described by attributes  $X_1, \dots, X_n$ . Potential alternatives are elements  $x = (x_1, \dots, x_n)$  of the Cartesian product  $X_1 \times \dots \times X_n =: X$ . We denote by  $N = \{1, \dots, n\}$  the index set of the attributes, and suppose throughout the paper

that  $n \geq 2$ . We employ the usual notation for compound alternatives, that is, for any  $x, y \in X$  and  $A \subseteq N$ , by  $(x_A, y_{-A})$  we mean the alternative taking value  $x_i$  for  $i \in A$  and  $y_i$  otherwise. We write  $x_{-i}$  instead of  $x_{-\{i\}}$  and extend this notation to the attributes as well:  $X_A, X_{-A}$ , etc.

The preference of the decision maker (DM) is represented by a binary relation  $\succsim$  on  $X$ , supposed to be complete and transitive. Ordinal measurement amounts to finding a numerical representation  $U : X \rightarrow \mathbb{R}$  of the preference in the sense that  $x \succsim y$  is equivalent to  $U(x) \geq U(y)$ .  $U$  is called a *value function*. A classical and simple example of value function is the *additive value function*

$$U(x) = \sum_{i=1}^n u_i(x_i) \quad (1)$$

where  $u_i : X_i \rightarrow \mathbb{R}$  are the *marginal value functions* on each attribute. It is well known that this model forces  $\succsim$  to satisfy *mutual preferential independence*: for any  $\emptyset \neq A \subset N$ , any  $x, y, z, t \in X$ ,

$$(x_A, z_{-A}) \succsim (y_A, z_{-A}) \Leftrightarrow (x_A, t_{-A}) \succsim (y_A, t_{-A}).$$

This strong condition is rarely met in practice, and usually one assume a much weaker version, where preferential independence is required only for singletons. Specifically,  $\succsim$  is said to satisfy *weak preferential independence* if for every  $i \in N$ , every  $x, y, z, t \in X$ ,

$$(x_i, z_{-i}) \succsim (y_i, z_{-i}) \iff (x_i, t_{-i}) \succsim (y_i, t_{-i}).$$

Under this condition, it is meaningful to define a preference relation  $\succsim_i$  over each attribute  $X_i$  as follows: for any  $x_i, y_i \in X_i$ ,  $x_i \succsim_i y_i$  if  $(x_i, z_{-i}) \succsim (y_i, z_{-i})$  for some  $z \in X$ . Then  $\succsim_i$  is a complete, transitive binary relation on  $X_i$ . It is easy to check that *monotonicity* holds:

$$x_i \succsim_i y_i \forall i \in N \Rightarrow x \succsim y. \quad (2)$$

## 2.2 Generalized Additive Independence (GAI) models

The additive value function model being too restrictive, one must look towards more general models. The *Generalized Additive Independence* model, proposed first by Fishburn (1967) (see also the pioneering work of Bacchus and Grove (1995)), is a natural generalization where the monodimensional marginal value functions are replaced by multidimensional marginals:

$$U(x) = \sum_{S \in \mathcal{S}} u_S(x_S) \quad (x \in X), \quad (3)$$

where  $\mathcal{S} \subseteq 2^N \setminus \{\emptyset\}$ . The additive value function model is recovered with  $\mathcal{S} = \{\{1\}, \dots, \{n\}\}$ . There is no specific requirement on the collection  $\mathcal{S}$  (hence  $S, T \in \mathcal{S}$  may overlap), nor on the marginal terms  $u_S$ .

The GAI model is very versatile, and may even violate the weak preference independence property. However, if we suppose that this property is true, (2) implies that a GAI model satisfying weak preferential independence is monotone:

$$x_i \succ_i y_i \forall i \in N \Rightarrow U(x) \geq U(y) \quad (4)$$

### 2.3 Discrete GAI models

We suppose from now on that the attributes take a finite number of values (discrete attributes):

$$X_i = \{a_i^0, \dots, a_i^{k_i}\} \quad (i \in N),$$

supposing that  $a_i^0 \preceq_i \dots \preceq_i a_i^{k_i}$ . Recall that under weak independence, the binary relations  $\preceq_i$ ,  $i = 1, \dots, n$  are complete preorders, and monotonicity (4) holds. For the sake of convenience, we normalize  $U$  by letting

$$U(a_1^0, \dots, a_n^0) = 0, \quad U(a_1^{k_1}, \dots, a_n^{k_n}) = 1. \quad (5)$$

Let us now simplify the notation. We replace each value  $a_i^\ell$  of attribute  $X_i$  by simply its index  $\ell$ . Doing so, an alternative  $(a_1^{\ell_1}, \dots, a_n^{\ell_n})$  is represented by  $(\ell_1, \dots, \ell_n)$  without ambiguity. Letting  $L := L_1 \times \dots \times L_n$ , with  $L_i = \{0, 1, \dots, k_i\}$ ,  $i \in N$ , this amounts to defining a bijection  $\varphi : X \rightarrow L$  with  $\varphi(a_1^{\ell_1}, \dots, a_n^{\ell_n}) = (\ell_1, \dots, \ell_n)$ . Thanks to the ordering  $a_i^0 \preceq_i \dots \preceq_i a_i^{k_i}$ ,  $v := U \circ \varphi^{-1} : L \rightarrow \mathbb{R}$  is a monotone function, which by (5) satisfies

$$v(0_N) = 0, v(k_N) = 1,$$

letting  $0_N = (0, \dots, 0)$ ,  $k_N = (k_1, \dots, k_n)$ .

From now on, we assume for simplicity that  $k_1 = k_2 = \dots = k_n =: k$  (this is without loss of generality, as the results presented hereafter remain valid for the general case). Such functions  $v$  are nothing other than  $k$ -ary capacities (Grabisch and Labreuche, 2003), which are particular *multichoice games* (Hsiao and Raghavan, 1990): a multichoice game is a function  $v : L \rightarrow \mathbb{R}$  satisfying  $v(0_N) = 0$ , and hence does not necessarily fulfill monotonicity. We denote by  $\mathcal{G}(L)$  the set of multichoice games defined on  $L$ , and by  $\mathcal{G}_M(L)$  the set of monotone multichoice games.

To summarize, we have considered a particular class of GAI models, namely those satisfying weak preferential independence and having discrete attributes with  $k$  values. Under these assumptions, the GAI model is equivalent to a  $k$ -ary capacity.

In Labreuche and Grabisch (to appear), the authors have considered continuous GAI models with the same assumption of weak preferential independence, by means of interpolation methods (Choquet integral, multilinear model) applied on a discrete model.

## 2.4 Models based on multichoice games

Thanks to multichoice games, we can be more general and drop the assumption of weak preferential independence, while keeping discrete attributes. Indeed, let us consider as above  $L = L_1 \times \dots \times L_n$  to be the set of alternatives (up to the mapping  $\varphi$ ), with  $\ell \in L_i$  corresponding to some value  $a_i^\ell$  of attribute  $X_i$ . As weak preferential independence does not hold any more, we cannot define an order on each attribute  $X_i$ . As a consequence, the function  $v = U \circ \varphi^{-1}$  is no more monotone and therefore is not a  $k$ -ary capacity but merely a multichoice game. Let us give a simple example borrowed from (Ridaoui et al., 2017a) to show that the situation is not so uncommon.

*Example 1.* The level of comfort of humans depends on three main attributes: temperature of the air ( $X_1$ ), humidity of the air ( $X_2$ ) and velocity of the air ( $X_3$ ). Then  $v(x_1, x_2, x_3)$  measures the comfort level. One can readily see that  $v$  is not monotone in its three arguments. For  $x_2$  and  $x_3$  fixed,  $v$  is maximal for intermediate values of the temperature (typically around 23°C). Similarly, the value of humidity maximizing  $v$  is neither too low nor too high. Finally, for  $x_1$  relatively large, some wind is well appreciated, but not too much. Hence for any  $i$ , and supposing the other two attributes being fixed, there exists an optimal value  $\hat{\ell}_i \in L_i$  such that  $v$  is increasing in  $x_i$  below  $\hat{\ell}_i$ , and then decreasing in  $x_i$  above  $\hat{\ell}_i$ .

Lack of monotonicity makes the analysis difficult. We will see in Section 3 how to define an importance index of attributes, which is valid for nonmonotonic models.

## 2.5 $p$ -additive models

$k$ -ary capacities are generalization of capacities introduced by Choquet (Choquet, 1953), while multichoice games generalize transferable utility (TU) games, introduced by Von Neumann and Morgenstern (von Neumann and Morgenstern, 1947). In these classical notions, we have  $k = 1$ , which amounts to considering set functions  $v : 2^N \rightarrow \mathbb{R}$ , with  $v(\emptyset) = 0$  (TU-games), and being monotone and satisfying  $v(N) = 1$  (capacities). For these functions, an important notion is the Möbius transform (Rota, 1964), which permits to define  $p$ -additive games and capacities (Grabisch, 1997). We introduce below these notions for the general case.

The *Möbius transform* of a multichoice game  $v$  is a function  $m^v : L \rightarrow \mathbb{R}$  which is the unique solution of the linear system

$$v(z) = \sum_{y \leq z} m^v(y) \quad (z \in L). \quad (6)$$

Its solution is shown to be (Grabisch and Labreuche, to appear)

$$m^v(z) = \sum_{y \leq z : z_i - y_i \leq 1 \forall i \in N} (-1)^{\sum_{i \in N} (z_i - y_i)} v(y) \quad (z \in \{0, 1, \dots, k\}^N). \quad (7)$$

It follows that any multichoice game  $v$  can be written as:

$$v = \sum_{x \in L^N \setminus \{0_N\}} m^v(x) u_x, \quad (8)$$

with  $u_x$  a  $k$ -ary capacity defined by

$$u_x(z) = \begin{cases} 1, & \text{if } z \geq x \\ 0, & \text{otherwise.} \end{cases}$$

By analogy with classical games,  $u_x$  is called the *unanimity game* centred on  $x$ . Note that this decomposition is unique as the unanimity games are linearly independent, and form a basis of the vector space  $\mathcal{G}(L)$ . Another basis is given by the Dirac games  $\delta_x$ , with  $x \in L$ ,  $x \neq 0_N$ :

$$\delta_x(y) = \begin{cases} 1, & \text{if } y = x \\ 0, & \text{otherwise.} \end{cases}$$

For further reference, let us introduce  $\mathcal{G}_+(L)$  the set of games with a nonnegative Möbius transform. Then any game  $v \in \mathcal{G}(L)$  can be expressed in a unique way as

$$v = v^+ - v^- \quad (9)$$

with  $v^+, v^- \in \mathcal{G}_+(L)$ .

We say that a multichoice game  $v$  is (*at most*)  $p$ -*additive* for some  $p \in \{1, \dots, n\}$  if its Möbius transform satisfies  $m^v(z) = 0$  whenever  $|\text{supp}(z)| > p$ , where

$$\text{supp}(z) = \{i \in N \mid z_i > 0\}.$$

The following result is shown in (Grabisch and Labreuche, to appear).

**Lemma 1.** *Let  $p \in \{1, \dots, n\}$ . A multichoice game  $v$  is  $p$ -additive if and only if it has the form*

$$v(z) = \sum_{x \in L, 0 < |\text{supp}(x)| \leq p} v_x(x \wedge z) \quad (z \in L) \quad (10)$$

where  $v_x : L \rightarrow \mathbb{R}$  with  $v_x(0_N) = 0$ .

Supposing that  $v$  is monotone, thanks to the bijection  $\varphi$ , the above result says that a discrete GAI model  $U$  is  $p$ -additive if and only if each term  $u_S$ ,  $S \in \mathcal{S}$ , has at most  $p$  variables, i.e.,  $|S| \leq p$ . In Section 4, we will study 2-additive GAI models. These models are of particular interest, because although they are much more general than the additive value function model, they remain tractable since each term depends of at most two variables.

### 3 Importance indices for discrete multicriteria decision models

We suppose in this section to have a MCDA model which is a multichoice game  $v$  on  $L$  (see Section 2.4).

The first level of interpretation of a MCDA model is to indicate to the user which attributes are the most important or influential in the decision model. This amounts to computing *importance indices* of each criterion in the model. **The knowledge of these values is very important. First, criteria of small importance index can be neglected. Second, the decision maker can rank the criteria by increasing importance according to his expertise. The comparison of this order with the order obtained from the importance indices is very informative. When there is some discrepancy, this means that the model have been underspecified, or there are some misunderstanding. The elicitation of the model has then to be updated.**

The MCDA models we are interested in can represent very rich and diverse decision strategies. The deciphering of complex MCDA models cannot be done from the sole knowledge of importance indices. One also needs information about how criteria interact together. We do not describe interaction indices in this chapter. We recommend reference (Ridaoui et al., 2018) to the interested reader.

This section is based on (Ridaoui et al., 2017a) and (Ridaoui et al., 2017b). We start by introducing the general idea of how to define an importance index (Section 3.1). Importance indices are closely related to the concept of a *value* in cooperative game theory. The existing literature in this field is summarized in Section 3.2. We consider two classes of importances indices. Section 3.3 defines the importance of criterion  $i$  as the average added-value (over all possible situations) of making a unitary improvement on criterion  $i$ . The sign of this index represents the general monotonicity of the model:  $U$  is globally nondecreasing (resp. nonincreasing) if the index is nonnegative (resp. nonpositive). This index is thus called *signed importance index*. It might happen that a function  $U$  that is nonincreasing in some area and nondecreasing in another area has an overall signed importance close to 0, even though this criterion is very important. We have thus also defined an *absolute importance index* (Section 3.4) to measure the net contribution of a criterion regardless of the monotonicity. Finally, we construct a very general class of importance indices written as a norm over all possible unitary improvement on a criterion (Section 3.5).

#### 3.1 How to define importance indices

The main ingredient behind importance indices, as for example it is defined for continuous functions of several variables (see, e.g., Grabisch et al. (2009)[§10.3]), is the average variation induced by a given variable over its domain, or equivalently, the average of the partial derivative of the function over its domain. We propose as a starting point to take this approach and to adapt it to  $v$ , which is a function defined over the discrete domain  $L$ . To this end, we introduce its *derivative w.r.t.  $i$* ,  $i \in N$ :

$$\Delta_i v(x) = v(x + 1_i) - v(x) \quad (x \in L, x_i < k),$$

where  $1_i$  is a shorthand for  $(1_i, 0_{-i})$ . Following the foregoing discussion, the general form of the importance index of attribute  $X_i$  w.r.t.  $v$  should read:

$$\phi_i(v) = \sum_{\substack{x \in L \\ x_i < k}} p_x^i \Delta_i v(x), \quad (11)$$

where  $p_x^i$  is a real positive constant, for every such  $x$  and  $i$ . If the weights  $p_x^i$  depend only on  $p_{x_{-i}}^i$ , then

$$\sum_{x_i=0}^{k_i-1} p_{x_{-i}}^i \Delta_i v(x) = p_{x_{-i}}^i (v(x_{-i}, k_i) - v(x_{-i}, 0_i)), \quad (12)$$

i.e., only the variation between  $k_i$  and  $0_i$  matters.

Clearly, if  $v$  is a monotone function, then  $\phi_i(v)$  is a nonnegative quantity for every  $i \in N$ , while it is a nonpositive quantity if  $v$  is antimonotonic. For this reason, one may call  $\phi(v)$  a *signed importance index*. What about nonmonotonic models, as the one given in Example 1? Taking 23°C as the optimal temperature, the derivative of  $v$  w.r.t. the temperature is positive for temperatures below 23°C, and becomes negative above 23°C. As a consequence, positive and negative variations may cancel each other, resulting in an importance index for temperature which is close to 0. This is quite counterintuitive, as surely temperature matters in the evaluation of comfort.

The foregoing discussion shows that a (signed) importance index is not adequate in any situation. Therefore, another definition seems to be necessary. The simplest idea to avoid this drawback is to cumulate the magnitude of the variations, regardless of their sign, instead of summing them algebraically. This leads to the following formula:

$$\phi_i(v) = \sum_{\substack{x \in L \\ x_i < k}} p_x^i |\Delta_i v(x)|. \quad (13)$$

We call such an index an *absolute importance index*. It coincides with the former one for monotonic games.

At this point of the discussion, one may say that only absolute importance indices are relevant, the signed ones leading to counterintuitive results. We think, however, that both are useful and should be used, provided we are aware of its precise meaning: The signed index indicates the overall trend of the model w.r.t. an attribute (increasing or decreasing), while the absolute index measures the amount of variation of the model induced by an attribute.

Finally, still other definitions can be proposed if one remarks that the absolute value is the  $L_1$  norm, and other norms can be used as well. We define *norm-based importance indices* as those of the following form:

$$\phi_i(v) = \sum_{\substack{x \in L \\ x_i < k}} p_x^i \|\Delta_i v(x)\|. \quad (14)$$

### 3.2 Values in game theory

We indicate in this section connections with cooperative game theory. In this field, a central notion is the one of value. Let us take for simplicity the case of classical TU-games. Interpreting  $v(N)$  as the total benefit achieved by the cooperation of all players, a *value* is a way of sharing  $v(N)$  among all the players, taking into account their contribution to the game. Formally, it is a mapping  $\phi : 2^N \rightarrow \mathbb{R}^N$ , and the usual requirement is that the value is *efficient*, which means that the benefit  $v(N)$  is shared without waste and nothing more can be given:  $\sum_{i \in N} \phi_i(v) = v(N)$ . The best known value is the Shapley value (Shapley, 1953), defined by:

$$\phi_i^{Sh}(v) = \sum_{S \subseteq N \setminus i} \frac{(n-s-1)!s!}{n!} (v(S \cup i) - v(S)), \forall i \in N. \quad (15)$$

Letting  $k = 1$ , the derivative becomes  $\Delta_i v(S) = v(S \cup i) - v(S)$ , hence the Shapley value has the form of a signed importance index. Indeed, in voting games, it is used as a power index, which is the counterpart of importance index for voting problems.

The Shapley value has been generalized to multichoice games, in different ways by several authors. We mention here (Hsiao and Raghavan, 1993) (historically the first one), van den Nouweland et al. (1995), Klijn et al. (1999), Peters and Zank (2005), Grabisch and Labreuche (2008), etc. The value of Peters and Zank reads:

$$\phi_i^{PZ}(v) = \sum_{x_{-i} \in \Gamma(L_{-i})} \frac{(n - \kappa(x_{-i}) - 1)! \kappa(x_{-i})!}{n!} (v(x_{-i}, k) - v(x_{-i}, 0)), \quad (16)$$

where  $\kappa(x_S)$  is the size of the kernel of  $x_S$ , i.e.,  $\kappa(x_S) = |\{i \in S \mid x_i = k\}|$ , and  $\Gamma(L_S) = \{0, k\}^S$ . Note that only vertices of  $L_{-i}$  are used in the computation.

All these values satisfy efficiency, as this is a basic requirement in a cooperation context. However, in a MCDA context, efficiency is not a relevant notion. Especially when  $v$  is not monotone, satisfying efficiency would lead to strange results. Indeed, as in Example 1, the value  $v(k_N)$  is low and even close to 0 since if all three parameters (temperature, humidity, wind velocity) take their maximal values, there is no comfort at all. As a consequence, the sum of the importance indices would be close to zero, and if absolute importance indices are taken (which are nonnegative by definition), the conclusion is that all criteria have a negligible importance, which is again quite counterintuitive.

As a conclusion, efficiency as defined in game theory must be abandoned, and none of the values defined in the literature can be taken as an importance index. In what follows, we propose and axiomatize different importance indices which are suitable for our MCDA context.

### 3.3 Signed importance indices

ction is to axiomatize the family of signed importance indices given by (11) and to propose a particular one based on suitable axioms.

The three first axioms we propose are the classical axioms used in the original axiomatization of Shapley. The first one says that  $\phi$  is a linear operator on  $\mathcal{G}(L)$ .

**Linearity axiom (L)** :  $\phi$  is linear on  $\mathcal{G}(L)$ , i.e., for any  $v, w \in \mathcal{G}(L)$ ,  $\forall \alpha \in \mathbb{R}$ ,

$$\phi_i(v + \alpha w) = \phi_i(v) + \alpha \phi_i(w).$$

An attribute  $i \in N$  is said to be null for  $v \in \mathcal{G}(L)$  if

$$v(x + 1_i) = v(x), \forall x \in L, x_i < k.$$

**Null axiom (N)**: If an attribute  $i$  is null for  $v \in \mathcal{G}(L)$ , then  $\phi_i(v) = 0$ .

This axiom says that an attribute for which an increment of 1 does not improve the evaluation is not important. It turns out that these two axioms are characteristic of the family of signed importance indices.

**Proposition 1.** *Under axioms (L) and (N), for all  $i \in N$ , there exist  $p_x^i \in \mathbb{R}$ , for all  $x \in L$  with  $x_i < k_i$ , such that for all  $v \in \mathcal{G}(L)$ ,*

$$\phi_i(v) = \sum_{\substack{x \in L \\ x_i < k_i}} p_x^i (v(x + 1_i) - v(x)). \quad (17)$$

We try now to refine the family by adding suitable properties. The first one is related to symmetry or anonymity: the numbering of the attributes should have no influence on the computation of the importance index.

Let  $\sigma$  be a permutation on  $N$ . For all  $x \in L$ , we denote  $\sigma(x)_{\sigma(i)} = x_i$ . For all  $v \in \mathcal{G}(L)$ , the game  $\sigma \circ v$  is defined by  $\sigma \circ v(\sigma(x)) = v(x)$ .

**Symmetry axiom (S)**: For any permutation  $\sigma$  on  $N$ ,  $\phi_{\sigma(i)}(\sigma \circ v) = \phi_i(v)$ ,  $\forall i \in N$ .

The next axiom is an invariance property. It says that the calculus of the importance index does not depend on the position on the “grid”  $L$ . It is another kind of symmetry axiom, relative to the levels  $0, 1, \dots, k$ , not to the attributes.

**Invariance axiom (I)**: Let us consider two games  $v, w \in \mathcal{G}(L)$  such that, for some  $i \in N$ ,

$$\begin{aligned} v(x + 1_i) - v(x) &= w(x) - w(x - 1_i), \forall x \in L, x_i \notin \{0, k\} \\ v(x_{-i}, 1_i) - v(x_{-i}, 0_i) &= w(x_{-i}, k_i) - w(x_{-i}, k_i - 1), \forall x_{-i} \in L_{-i}. \end{aligned}$$

Then  $\phi_i(v) = \phi_i(w)$ .

With these two additional axioms, we obtain the following result.

**Proposition 2.** *Under axioms (L), (N), (I) and (S), for all  $v \in \mathcal{G}(L)$ , for all  $i \in N$ ,*

$$\phi_i(v) = \sum_{x_{-i} \in L_{-i}} p_{n(x_{-i})} (v(x_{-i}, k_i) - v(x_{-i}, 0_i)),$$

where  $n(x_{-i}) = (n_0, n_1, \dots, n_k)$  with  $n_j$  the number of components of  $x_{-i}$  being equal to  $j$ .

The effect of the two axioms is the following: **(I)** forces  $p_x^i$  to depend only on  $x_{-i}$  and  $i$ , which by (12) implies that only the difference between  $k_i$  and  $0_i$  on attribute  $i$  matters. Then the symmetry axiom makes the constant  $p_{x_{-i}}^i$  to depend only on the “cardinality” of  $x_{-i}$ .

It remains to find a last axiom for determining the constants uniquely. As the usual efficiency axiom of game theory is not suitable in this context, we propose instead a substitute which is in the spirit of variation calculus:

**Efficiency axiom (E):** For all  $v \in \mathcal{G}(L)$ ,

$$\sum_{i \in N} \phi_i(v) = \sum_{\substack{x \in L \\ x_j < k}} (v(x + 1_N) - v(x)).$$

It can be explained as follows: taking an alternative  $x \in L$  and increasing the value of each attribute by one unit, i.e., going to  $x + 1_N$ , the amount of variation is due to the contribution of all attributes, and the sum of all importance indices should be equal to the sum of this variation for all alternatives  $x$ . Interestingly, the axiom is nevertheless not so far from the original efficiency axiom because when taking  $k = 1$ , it reduces to the classical efficiency axiom  $\sum_i \phi_i(v) = v(N)$ .

Finally, we can show:

**Theorem 1.** *Under axioms (L), (N), (I), (S) and (E), for all  $v \in \mathcal{G}(L)$*

$$\phi_i(v) = \phi_i^S(v) := \sum_{x_{-i} \in L_{-i}} \frac{(n - \sigma(x_{-i}) - 1)! \kappa(x_{-i})!}{(n + \kappa(x_{-i}) - \sigma(x_{-i}))!} (v(x_{-i}, k_i) - v(x_{-i}, 0_i)), \forall i \in N. \quad (18)$$

### 3.4 Absolute importance indices

We turn to the axiomatization of the family of absolute importance indices and as before try to find a particular index of interest.

The major difficulty in axiomatizing (13) is that  $\phi$  does not satisfy linearity. Therefore, it is not possible to start from the decomposition of a game on some basis. We remark that if  $v$  is monotone, then  $|v(x + 1_i) - v(x)| = v(x + 1_i) - v(x)$  for every  $x \in L, x_i < k$ . However,  $\mathcal{G}_M(L)$  is not a linear subspace of  $\mathcal{G}(L)$  but a convex

cone, and we cannot apply directly the linearity axiom on it. The idea is the following: using the expression of  $v$  in the basis of unanimity games (8), this expression turns to be a conic combination iff  $v$  is in  $\mathcal{G}_+(L)$ . As any game can be written as the difference of two games in  $\mathcal{G}_+(L)$  (see (9)), it is then possible to extend this expression to monotone games. Hence,  $\phi$  should commute with conic combination and differences of games in  $\mathcal{G}_+(L)$ .

**Conic Combination axiom (CC)** : For every  $v, w \in \mathcal{G}_+(L)$ , for every  $\alpha \in \mathbb{R}_+$ ,

$$\phi(v + \alpha w) = \phi(v) + \alpha \phi(w).$$

**Decomposition axiom (D)**: If  $v, v' \in \mathcal{G}_+(L)$  and  $v - v'$  is monotone, then  $\phi(v - v') = \phi(v) - \phi(v')$ .

These two axioms permit to obtain the following result.

**Proposition 3.** *Under axioms (CC) and (D), for all  $i \in N$ , there exists constants  $a_x^i \in \mathbb{R}$ , for all  $x \in L$ , such that  $\forall v \in \mathcal{G}_M(L)$ ,*

$$\phi_i(v) = \sum_{x \in L} a_x^i v(x). \quad (19)$$

Taking two multichoice games  $v$  and  $w$  for which the marginal contribution of an attribute  $i$  to a game  $v$  is the same or the opposite of that to a game  $w$ , the average importance of attribute  $i$  shall be the same for  $v$  and  $w$ .

**Marginal contribution axiom (MC)**: Let  $i \in N$  and  $v, w \in \mathcal{G}(L)$  such that

$$|\Delta_i(v)(x)| = |\Delta_i(w)(x)|, \forall x \in L, x_i < k.$$

Then

$$\phi_i(v) = \phi_i(w).$$

The following result shows that the family of absolute importance indices is characterized by the above three axioms.

**Proposition 4.** *Under axioms (CC), (D) and (MC), there exist real constants  $p_x^i$ ,  $i \in N, x \in L, x_i < k_i$ , such that for every  $v \in \mathcal{G}(L)$ ,*

$$\phi_i(v) = \sum_{\substack{x \in L \\ x_i < k_i}} p_x^i |\Delta_i v(x)|. \quad (20)$$

Surprisingly, there is no need of the null axiom. This is because it is implied by (MC) and (CC) (or by (MC) and (D)) as it is easy to check.

As for the signed importance index, the introduction of the two symmetry axioms (S) and (I) permits to reduce the number of constants, as  $p_x^i$  is turned into  $p_{n(x_{-i})}$ , where  $n(x_{-i}) = (n_0, n_1, \dots, n_k)$  with  $n_j$  the number of components of  $x_{-i}$  being equal to  $j$ .

It remains to determine uniquely the constants by imposing some normalization condition. The first one is based on the Dirac games  $\delta_x$ . Observe that if  $x_i \neq 0, k$ , the sum of absolute variations along the  $i$  axis is 2, otherwise it is 1. Normalizing by the total number of points in the grid  $L_{-i}$ , which is  $(k+1)^{n-1}$  so that the result is not dependent of the size of the grid, we obtain the following:

**Calibration axiom 1st version (C1):** For every  $x \in L \setminus \{0_N\}$

$$\phi_i(\delta_x) = \begin{cases} 2/(k+1)^{n-1} & \text{if } 0 < x_i < k \\ 1/(k+1)^{n-1} & \text{otherwise.} \end{cases}$$

**Theorem 2.** Under axioms (CC), (D), (S), (I), (MC) and (C1), for all  $v \in \mathcal{G}(L)$

$$\phi_i(v) = \frac{1}{(k+1)^{n-1}} \sum_{\substack{x \in L \\ x_i < k}} |v(x+1_i) - v(x)|, \forall i \in N,$$

Another possibility is based on unanimity games.

**Calibration axiom 2nd version (C2):** For all  $x \in L \setminus \{0_N\}$ , for all  $i \in S(x)$ ,

$$\phi_i(u_x) = \frac{1}{s(x)}.$$

**Theorem 3.** Under axioms (CC), (D), (S), (I), (MC) and (C2), for all  $v \in \mathcal{G}(L)$

$$\phi_i(v) = \sum_{\substack{x_{-i} \in \{0,k\}^{N \setminus \{i\}} \\ x_i \in L_i, x_i < k}} \frac{(n - s(x_{-i}) - 1)! s(x_{-i})!}{n!} |v(x+1_i) - v(x)|, \forall i \in N.$$

We observe that (C1) yields an importance index similar to the Banzhaf value, while the use of (C2) gives a result close to the Shapley value. It is possible, however, to obtain exactly the coefficients of the signed importance index (see Theorem 1) by using suitable axioms (Grabisch et al., submitted).

### 3.5 Norm-based importance indices

We turn now to the family of norm-based importance indices, which have the general form (5). A fundamental difficulty is that the use of a norm different from  $L_1$  forbids to take an axiomatic approach similar to the one we used for absolute importance indices, because there would exist no class of games where a property similar to linearity would hold. Nevertheless, it is possible to obtain a general form through a number of axioms which are presented below. In the rest of this section  $i \in N$  is fixed.

**Nonnegativity (NN)** : The importance index takes nonnegative values, i.e.,  
 $\phi_i : \mathcal{G}(L) \rightarrow \mathbb{R}_+$ .

**Absolute Homogeneity (AH)**: For every  $\alpha \in \mathbb{R}$  and every game  $v \in \mathcal{G}(L)$ ,

$$\phi_i(\alpha v) = |\alpha| \phi_i(v)$$

**Subadditivity (SA)** : For any games  $v, w \in \mathcal{G}(L)$ ,

$$\phi(v + w) \leq \phi(v) + \phi(w)$$

**Strong Null axiom (SN)** :  $\phi_i(v) = 0$  if and only if  $i$  is null for  $v$ .

The nonnegativity axiom says that importance indices are nonnegative quantities. Absolute homogeneity says that multiplying a game by a constant just multiplies the importance index by the magnitude of this constant. The subadditivity axiom expresses the fact that summing two games  $v, w$  may hinder the importance of an attribute by some hedging effect: the positive variation of  $i$  at some point  $x$  for  $v$  can be cancelled by a negative variation at the same point for  $w$ . Lastly, the strong null axiom is a strong version of the usual null axiom, in the sense that *only* games whose attribute  $i$  is null can lead to a null importance index for  $i$ .

We obtain the following.

**Theorem 4.** *Under axioms (NN), (AH), (SA) and (SN), there exists a norm  $\|\cdot\|$  on  $\mathbb{R}^{k(k+1)^{n-1}}$  and a linear one-to-one mapping  $h$  on  $\mathbb{R}^{k(k+1)^{n-1}}$  such that*

$$\phi_i(v) = \|h \circ \Delta_i(v)\|.$$

## 4 Monotone decomposition of a 2-additive GAI model and its interpretation

Following Section 3, a first level of interpretation of a MCDA model consists in using the generic indices, such as the importance indices. To go further, one needs to take profit of the particular form (in particular the mathematical expression) of utility  $U$ . We will focus in this section on the GAI model. As it takes an additive form, it would be convenient to interpret  $U$  by interpreting each term  $u_S$  separately. We will see that the decomposition of a GAI model is far from being unique, which makes its interpretation delicate. We focus on the special class of the 2-additive GAI models that are monotone, as monotonicity is a very natural property in MCDA. We will see whether monotonicity improves the interpretability of a GAI model.

### 4.1 Difficulty of the interpretation of a GAI model

We formalize in this section the intuition given earlier on the difficulty of interpreting a GAI model.

#### 4.1.1 Illustration of the difficulty on an example

A model very similar to the 2-additive GAI model is defined in (Greco et al., 2014). In this reference, the sign of interacting terms  $u_{i,j}$  is interpreted as the sign of interaction. This is borrowed from the expression of the Choquet integral with respect to a 2-additive capacity  $v$ , which takes the following form written in terms of the Möbius transform (Choquet, 1953):

$$\text{Ch}_v(a_1, \dots, a_n) = \sum_{i \in N} m^v(\{i\}) a_i + \sum_{\{i,j\} \subseteq N} m^v(\{i,j\}) a_i \wedge a_j,$$

where  $m^v$  is the Möbius transform of  $v$  (see Section 2.5). The interaction coefficient between factors  $i$  and  $j$  is equal  $m^v(\{i,j\})$ , so that the sign of the interaction is given by the sign of  $m^v(\{i,j\})$ . However, this interpretation holds only for the expression of the Choquet integral with respect to the Möbius transform, that is, the expression of  $\text{Ch}$  on the basis  $\{a_i, i \in N\} \cup \{a_i \wedge a_j, \{i,j\} \subseteq N\}$ . This is no more true for another basis. In order to illustrate this, let us take the following example.

*Example 2.* Let us take the following function of two variables:

$$U(x_1, x_2) = 2x_1 + x_2 - \min(x_1, x_2). \quad (21)$$

Following the intuition of (Greco et al., 2014), one would say that there is a negative interaction between the two attributes. However, using the relation  $\max(x_1, x_2) + \min(x_1, x_2) = x_1 + x_2$ , we obtain an equivalent expression

$$U(x_1, x_2) = x_1 + \max(x_1, x_2) \quad (22)$$

in which the bivariate term is now nonnegative.

Apart from the problem of the sign of the interaction, relations (21) and (22) are two different equivalent expressions of the same model  $U$ . If one wants to present the interacting term  $u_{i,j}$  to the user, which one among terms “ $-\min(x_1, x_2)$ ” and “ $\max(x_1, x_2)$ ” shall be shown?

#### 4.1.2 Nonuniqueness of the GAI decomposition

We have already seen that the GAI decomposition (3) is not unique. In the additive utility model, each utility term is given up to a constant. Comparing (21) and (22), we see that the terms in the GAI decomposition can take very different expressions.

This rises the question of the *decomposition* of a GAI model  $U(x)$ . Is it possible to relate all decompositions? Fishburn (1967) has shown that any two equivalent decompositions  $U(x) = \sum_{S \in \mathcal{S}} u_S(x_S) = \sum_{S \in \mathcal{S}} u'_S(x_S)$  are related as follows:

$$u'_S(x_S) = u_S(x_S) + \sum_{S' \in \mathcal{S} \setminus \{S\}, S \cap S' \neq \emptyset} f_{S,S'}(x_{S \cap S'}) + c_S \quad (23)$$

where  $f_{S,S'} : X_{S \cap S'} \rightarrow \mathbb{R}$ , and  $\sum_{S \in \mathcal{S}} [\sum_{S' \in \mathcal{S} \setminus \{S\}, S \cap S' \neq \emptyset} f_{S,S'}(x_{S \cap S'}) + c_S] = 0$ . There is an intrinsic difficulty coming from the nonuniqueness of the GAI decomposition. In the interpretation of model  $U$ , which specific decomposition shall be used? This raises the question of whether there exists a canonical decomposition, which would allow for an intuitive interpretation of  $U$ . This question will be addressed in Section 4.4.

## 4.2 Monotonicity conditions

Monotonicity is an essential property in MCDA. We assume that  $U$  satisfies monotonicity condition (2). If  $U$  is monotone and takes the form (3), it would be counter-intuitive and misleading for the end-user if we present him some terms  $u_S$  that are not monotone in some of their coordinates.

In Example 2, it is apparent from (22) that  $U$  is monotone in the two attributes. Expression (21) is formed of three terms, the first two being increasing while the third one is decreasing. Presenting this to the user would be confusing because of the decreasing term. On the other hand, (22) has only two terms, both of them being nondecreasing. Clearly, the latter expression is more transparent because it has fewer terms and each term in the expression is both nonnegative and nondecreasing.

We have seen that, in two equivalent expressions (such as (21) and (22)), a similar term  $u_{1,2}$  (“ $-\min(x_1, x_2)$ ” and “ $\max(x_1, x_2)$ ” respectively) does not have the same monotonicity. This cannot happen with the additive utility model and shows in particular that the GAI model does not necessarily satisfy weak independence. In the Artificial Intelligence community, researchers are interested in the representation of preferences that may violate weak independence. A well-known example of such a preference is the following: consider the choice of a menu described by two attributes  $X_1, X_2$  where  $X_1$  pertains on the type of wine and  $X_2$  to the type of main course in a restaurant. Then usually, one prefers ‘red wine’ to ‘white wine’ if the main course is ‘meat’, but ‘white wine’ is preferred to ‘red wine’ if the main course is ‘fish’ (the preference over attribute ‘wine’ is conditional on the value on attribute ‘main course’) (Boutilier et al., 2001). This can be represented by a GAI model:  $U(x_1, x_2) = u_{1,2}(x_1, x_2) + u_2(x_2)$ , where

- preferences over  $X_2$ :  $u_2(\text{meat}) = 1$ ,  $u_2(\text{fish}) = 0$ ,
- preferences over  $X_1$  conditionally on  $X_2$ :  $u_{1,2}(\text{red}, \text{meat}) = 4$ ,  $u_{1,2}(\text{white}, \text{meat}) = 2$ ,  $u_{1,2}(\text{white}, \text{fish}) = 3$  and  $u_{1,2}(\text{red}, \text{fish}) = 0$ .

The examples which violate weak independence are far from being the general case in MCDA. Rather, most of MCDA problems satisfy weak independence, as there are in general natural preferences on each attribute. For this reason, following the traditional view of decision theory, we assume in this chapter that weak independence holds.

An important consequence of weak independence is that monotonicity holds for  $\succsim$  (see (2)), and consequently for  $U$  too - see (4). We note that in (23), due to the presence of functions  $f_{S,S'}$ , we do not have  $u_S(x_S) \geq u_S(y_S)$  iff  $u'_S(x_S) \geq u'_S(y_S)$ , for any two  $x_S, y_S \in X_S$  (Braziunas, 2012, page 87). Moreover, even if  $U$  satisfies weak independence, it might be the case that  $u_S$  does not fulfil this condition, or satisfies it but does not have the same monotonicity as  $U$ .

### 4.3 Representation of monotone 2-additive GAI models

We have seen in Example 2 a situation where, starting from a monotone 2-additive GAI model (namely expression (21)), we can find an equivalent expression (namely (22)) such that each term is nonnegative and monotone. The main question we wish to address in this section is the following one: is the previous repair process working in all situations?

#### 4.3.1 Main result

The following theorem states that a decomposition of a 2-additive monotone GAI model into monotone nondecreasing terms is always possible.

**Theorem 5 ((Grabisch and Labreuche, to appear)).** *Let us consider a 2-additive discrete GAI model  $U$  satisfying monotonicity (4) and (5). Then there exist non-negative and nondecreasing functions  $u_i : X_i \rightarrow [0, 1]$ ,  $i \in N$ ,  $u_{ij} : X_i \times X_j \rightarrow [0, 1]$ ,  $\{i, j\} \subseteq N$ , such that*

$$U(x) = \sum_{i \in N} u_i(x_i) + \sum_{\{i, j\} \subseteq N} u_{ij}(x_i, x_j) \quad (x \in X) \quad (24)$$

The rest of Section 4.3 is devoted to describing important consequences of this important result.

#### 4.3.2 Consequence of the main result in the computation of the importance indices

Let us consider a 2-additive monotone GAI model  $U$  (3). The first level of interpretation of  $U$  aims at computing the importance indices and showing to the user the most important attributes. Let us start with the signed importance index  $\phi_i^s(U \circ \varphi^{-1})$

(see Section 3.3). The main drawback of formula (18) is that it has an exponential number of terms in the number  $n$  of criteria. Fortunately, we can drastically reduce this complexity for GAI models. Indeed, as  $\phi_i^s$  fulfills **(L)**, we can write

$$\phi_i^s(U \circ \varphi^{-1}) = \sum_{S \in \mathcal{S}} \phi_i^s(u_S \circ \varphi^{-1}).$$

For a 2-additive GAI model,  $S \in \mathcal{S}$  contains at most two elements, which makes the computation of each term  $\phi_i^s(u_S \circ \varphi^{-1})$  extremely fast. Hence the computation of  $\phi_i^s(U \circ \varphi^{-1})$  becomes easily tractable (in  $O(n^2)$ ) even for large values of  $n$ .

Two absolute importance indices have been proposed in Section 3.4. These two expressions have different coefficients compared to  $\phi_i^s$ . As we mentioned in Section 3.4, it is possible to define an absolute importance index having the same coefficient as in  $\phi_i^s$  (Grabisch et al., submitted):

$$\phi_i^a(v) := \sum_{x_{-i} \in L_{-i}} \frac{(n - \sigma(x_{-i}) - 1)! \kappa(x_{-i})!}{(n + \kappa(x_{-i}) - \sigma(x_{-i}))!} |v(x_{-i}, k_i) - v(x_{-i}, 0_i)|. \quad (25)$$

Then if  $U$  is monotone, we have

$$\phi_i^s(U \circ \varphi^{-1}) = \phi_i^a(U \circ \varphi^{-1}).$$

On the other hand, nothing a priori forbids each term  $u_S$  in the GAI decomposition to be nonpositive or nonmonotonic. In this case one would have

$$\phi_i^s(u_S \circ \varphi^{-1}) \neq \phi_i^a(u_S \circ \varphi^{-1}).$$

This would be counter-intuitive and misleading for the user. Fortunately, according to Theorem 5, there exists a GAI decomposition in which each term  $u_S$  is non negative and monotone. For such decomposition, we obtain  $\phi_i^s(u_S \circ \varphi^{-1}) = \phi_i^a(u_S \circ \varphi^{-1})$ .

### 4.3.3 A complexity problem in the learning procedure

(see (Grabisch and Labreuche, to appear) for more details) Another very important consequence of Theorem 5 concerns the learning of the GAI model. Generally speaking, a learning procedure consists in collecting information on the preference relation  $\succsim$ , which is then used in an optimization problem, whose aim is to find a value function  $U$  of a given type representing at best the given preference. The variables of the optimization problem are then the parameters of the model  $U$ .

We begin by computing the number of unknowns in a 2-additive GAI model equivalent to a  $k$ -ary capacity. Such a model has the form (3) with  $\mathcal{S}$  being the set of singletons and pairs. Since  $|L_i| = k + 1$ , this yields

$$\eta(k, n) = (k + 1) \binom{n}{1} + (k + 1)^2 \binom{n}{2} = \frac{n(k + 1)}{2} (2 + (k + 1)(n - 1))$$

unknowns.  $U$  being monotone nondecreasing, this induces a number of monotonicity constraints on the unknowns, of the type

$$U(a_1^{j_1}, \dots, a_{i-1}^{j_{i-1}}, a_i^{j_i+1}, a_{i+1}^{j_{i+1}}, \dots, a_n^{j_n}) \geq U(a_1^{j_1}, \dots, a_{i-1}^{j_{i-1}}, a_i^{j_i}, a_{i+1}^{j_{i+1}}, \dots, a_n^{j_n}) \quad (26)$$

for every  $i \in N$ ,  $j_1 \in \{0, \dots, k_1\}, \dots, j_{i-1} \in \{0, \dots, k_{i-1}\}, j_i \in \{0, \dots, k_i - 1\}, j_{i+1} \in \{0, \dots, k_{i+1}\}, \dots, j_n \in \{0, \dots, k_n\}$ . The number of elementary conditions contained in (26) is equal to

$$\sum_{i \in N} \left( k_i \times \prod_{j \in N \setminus \{i\}} (k_j + 1) \right).$$

In the case where  $k_i = k$  for every  $i$ , this number becomes

$$\kappa(k, n) = n \times k \times (k + 1)^{n-1}.$$

Although the number of variables  $\eta(k, n)$  is still quadratic in  $n$  and  $k$ , the number of constraints  $\kappa(k, n)$  is exponential in  $n$ . It follows that any practical identification of a GAI model based on some optimization procedure<sup>1</sup>, where the variables are the unknowns of the GAI model and the constraints are the monotonicity constraints (26) plus possibly some learning data, has to cope with an exponential number of constraints. The following tables, obtained with  $k = 4$ , shows that the underlying optimization problem becomes rapidly intractable.

$n$	4	6	8	10
$\eta(4, n)$	170	405	740	1175
$\kappa(4, n)$	2000	75 000	2 500 000	78 125 000

  

$n$	12	14	20
$\eta(k, n)$	1710	2345	4850
$\kappa(k, n)$	2 343 750 000	68 359 375 000	1.526E + 15

However, if a decomposition into nonnegative nondecreasing terms is possible, one has only to check monotonicity of each term:

$$\forall i \in N \forall l \in \{0, \dots, k_i - 1\} \quad u_i(a_i^{l+1}) \geq u_i(a_i^l), \quad (27)$$

$$\forall \{i, j\} \subseteq N \forall l_i \in \{0, \dots, k_i - 1\} \forall l_j \in \{0, \dots, k_j\} \\ u_{i,j}(a_i^{l_i+1}, a_j^{l_j}) \geq u_{i,j}(a_i^{l_i}, a_j^{l_j}), \quad (28)$$

$$\forall \{i, j\} \subseteq N \forall l_i \in \{0, \dots, k_i\} \forall l_j \in \{0, \dots, k_j - 1\} \\ u_{i,j}(a_i^{l_i}, a_j^{l_j+1}) \geq u_{i,j}(a_i^{l_i}, a_j^{l_j}). \quad (29)$$

Then the number of monotonicity conditions drops to

<sup>1</sup> The learning problem can be classically transformed into a linear program, where the training set is seen as linear constraints on the GAI variables (Bigot et al., 2012; Greco et al., 2014). It could also be possible to perform statistical learning, like in Fallah Tehrani et al. (2012), where the underlying optimization problem is a convex problem under linear constraints.

$$\sum_{i \in N} k_i + \sum_{\{i,j\} \subseteq N} (k_i(k_j + 1) + k_j(k_i + 1)).$$

In the case where  $k_i = k$  for every  $i$ , this number becomes

$$\kappa'(k, n) = n \times k \times [(n-1)(k+1) + 1],$$

which is quadratic in  $n$ . The following table ( $k = 4$ ) shows that the optimization problem becomes tractable even for a large number of attributes.

$n$	4	6	8	10	12	14	20
$\kappa'(k, n)$	256	624	1152	1840	2688	3696	7680

#### 4.4 Interpretation through a canonical decomposition

The most convenient way to interpret a GAI model is to use a “canonical” decomposition in some sense. Following Theorem 5, is it possible to always find decomposition into nonnegative and nondecreasing terms? We review in this section two existing decompositions of a multivariable function into a GAI decomposition. We conclude by providing some research directions to construct a canonical decomposition of a GAI model fulfilling the monotonicity of each of its terms.

##### 4.4.1 Braziunas’ Decomposition

Braziunas has proposed a decomposition based on the Fishburn representation (Braziunas, 2012). Fixing an order on  $\mathcal{S}$ , say,  $\mathcal{S} = \{S_1, \dots, S_p\}$ , the overall value function reads  $U(x) = \sum_{S \in \mathcal{S}} u_S^C(x_S)$  with, for every  $j \in \{1, \dots, p\}$

$$u_{S_j}^C(x_{S_j}) = U(x[S_j]) + \sum_{K \subseteq \{1, \dots, j-1\}, K \neq \emptyset} (-1)^{|K|} U(x[\cap_{k \in K} S_k \cap S_j]) \quad (30)$$

where  $\cdot^C$  stands for “canonical”,  $\mathbb{O} \in X$  is any element in  $X$  seen as an anchor, and  $x[S] \in X$  defined by  $(x[S])_i = x_i$  if  $i \in S$  and  $(x[S])_i = \mathbb{O}_i$  otherwise (Braziunas, 2012, page 94)). Note that the expression depends on the chosen ordering of the elements of  $\mathcal{S}$ . The two equivalent decompositions (21) and (22) were obtained with a particularly simple example. The previous remark provides a more systematic way to derive several equivalent decompositions of GAI models, as illustrated in the next example.

*Example 3.* Consider the following function  $U(x_1, x_2, x_3) = x_2 + x_1 x_3 + \max(x_1, x_2)$ . We have  $\mathcal{S} = \{S_1, S_2, S_3\}$  with  $S_1 = \{2\}$ ,  $S_2 = \{1, 3\}$  and  $S_3 = \{1, 2\}$ . Then the canonical decomposition gives, with  $\mathbb{O} = (0, 0, 0)$ :

$$\begin{aligned}
u_{S_1}^C(x_2) &= U(x[S_1]) = U(\mathbb{O}_1, x_2, \mathbb{O}_3) = 2x_2 \\
u_{S_2}^C(x_1, x_3) &= U(x[S_2]) - U(x[S_1 \cap S_2]) = U(x_1, \mathbb{O}_2, x_3) - U(\mathbb{O}) = x_1(x_3 + 1) \\
u_{S_3}^C(x_1, x_2) &= U(x[S_3]) - U(x[S_1 \cap S_3]) - U(x[S_2 \cap S_3]) + U(x[S_1 \cap S_2 \cap S_3]) \\
&= U(x_1, x_2, \mathbb{O}_3) - U(\mathbb{O}_1, x_2, \mathbb{O}_3) - U(x_1, \mathbb{O}_2, \mathbb{O}_3) + U(\mathbb{O}) \\
&= \max(x_1, x_2) - x_1 - x_2 = -\min(x_1, x_2)
\end{aligned}$$

We note that  $U$  is nondecreasing in all variables, even though, for the canonical decomposition,  $u_{S_3}^C$  is nonincreasing in its two coordinates.

Let us take now the order  $S'_1 = \{1, 2\}$ ,  $S'_2 = \{1, 3\}$  and  $S'_3 = \{2\}$ . We obtain  $u_{S'_1}^C(x_1, x_2) = x_2 + \max(x_1, x_2)$ ,  $u_{S'_2}^C(x_1, x_3) = x_1 x_3$  and  $u_{S'_3}^C(x_2) = 0$ . All terms are now nonnegative and monotone.

The previous example shows that the canonical decomposition does not guarantee to have only nondecreasing terms in the decomposition, and therefore does not provide an easily interpretable decomposition. Hence there is no well-defined semantics of the value functions  $u_S$ , contrarily to what is claimed in (Braziunas, 2012, section 3.2.1.4).

#### 4.4.2 ANOVA

We present in this section ANOVA (Fisher and Mackenzie, 1923), which is a convenient way to construct a canonical decomposition of a GAI model.

In statistics, the analysis of variance (ANOVA) is a powerful tool to represent interaction between variables in a multivariate model (Fisher and Mackenzie, 1923). Consider  $n$  independent random variables  $Z_1, \dots, Z_n$  uniformly distributed in  $[0, 1]$ , and a multivariate model  $Y = F(Z)$ , where  $Z = (Z_1, \dots, Z_n)$ . Let us denote by  $Z_S$  and  $Z_{-S}$  the groups of variables  $(Z_i)_{i \in S}$  and  $(Z_i)_{i \notin S}$  respectively. Hence, we may write  $Z = (Z_S, Z_{-S})$ . Moreover, we denote by  $\mathbb{E}[Y]$  the expected value of  $Y$  taken over all variables  $Z_1, \dots, Z_n$ . The expected value of  $Y$  can be taken on a subset  $Z_S$  of variables, with the corresponding notation  $\mathbb{E}_{Z_S}[Y]$ .

Any multivariate function can be decomposed in the following way (ANOVA decomposition) (Fisher and Mackenzie, 1923):

$$Y = F(Z) = F_{\emptyset} + \sum_{i=1}^n F_i(Z_i) + \sum_{i < j} F_{ij}(Z_i, Z_j) + \dots + F_N(Z) = \sum_{S \subseteq N} F_S(Z_S),$$

with

$$\begin{aligned}
F_\emptyset &= \mathbb{E}[Y] \\
F_i(Z_i) &= \mathbb{E}[Y|Z_i] - F_\emptyset \\
F_{ij}(Z_i, Z_j) &= \mathbb{E}[Y|Z_i, Z_j] - F_i(Z_i) - F_j(Z_j) - F_\emptyset \\
&= \mathbb{E}[Y|Z_i, Z_j] - \mathbb{E}[Y|Z_i] - \mathbb{E}[Y|Z_j] + E[Y] \\
&\vdots \\
F_S(Z_S) &= \mathbb{E}_{Z_{-S}}[Y|Z_S] - \sum_{T \subset S} F_T(Z_T) = \sum_{T \subset S} (-1)^{|S \setminus T|} \mathbb{E}_{Z_{-T}}[Y|Z_T] \\
&\vdots \\
F_N(Z) &= \sum_{T \subset N} (-1)^{|N \setminus T|} \mathbb{E}_{Z_{-T}}[Y|Z_T].
\end{aligned}$$

We note that the ANOVA decomposition corresponds exactly to a GAI decomposition. If we start with a function  $U$  taking a GAI decomposition with a collection  $\mathcal{S}$  of represented subsets, then one can easily show that the ANOVA decomposition applied to  $U$  cannot return nonzero terms  $U_S$  where  $S$  is a superset of some terms in  $\mathcal{S}$ . In other word, the ANOVA decomposition will use subsets in  $\widehat{\mathcal{S}} = \{S \subseteq S', S' \in \mathcal{S}\}$ .

Let us apply the ANOVA decomposition to the example (22):  $U(x_1, x_2) = x_1 + \max(x_1, x_2)$ . We have

$$\begin{aligned}
U_\emptyset &= \mathbb{E}[Y] = \int_0^1 \int_0^1 (x_1 + \max(x_1, x_2)) dx_1 dx_2 \\
&= \frac{1}{2} + \int_0^1 \int_0^{x_1} x_1 dx_1 dx_2 + \int_0^1 \int_{x_1}^1 x_2 dx_1 dx_2 \\
&= \frac{1}{2} + \int_0^1 x_1^2 dx_1 + \int_0^1 \frac{1-x_1^2}{2} dx_1 = \frac{1}{2} + \frac{1}{3} + \left(\frac{1}{2} - \frac{1}{6}\right) = \frac{7}{6}
\end{aligned}$$

$$\begin{aligned}
U_1(x_1) &= \mathbb{E}[U|x_1] - U_\emptyset = \int_0^1 (x_1 + \max(x_1, x_2)) dx_2 - \frac{7}{6} \\
&= x_1 + \int_0^{x_1} x_1 dx_2 + \int_{x_1}^1 x_2 dx_2 - \frac{7}{6} = x_1 + x_1^2 + \frac{1-x_1^2}{2} - \frac{7}{6} = x_1 + \frac{x_1^2}{2} - \frac{2}{3}
\end{aligned}$$

$$\begin{aligned}
U_2(x_2) &= \mathbb{E}[U|x_2] - U_\emptyset = \int_0^1 (x_1 + \max(x_1, x_2)) dx_1 - \frac{7}{6} \\
&= \frac{1}{2} + \int_0^{x_2} x_2 dx_1 + \int_{x_2}^1 x_1 dx_1 - \frac{7}{6} = -\frac{2}{3} + x_2^2 + \frac{1-x_2^2}{2} = \frac{x_2^2}{2} - \frac{1}{6}
\end{aligned}$$

$$\begin{aligned}
U_{12}(x_1, x_2) &= U(x_1, x_2) - U_1(x_1) - U_2(x_2) + U_\emptyset \\
&= (x_1 + \max(x_1, x_2)) - \left(x_1 + \frac{x_1^2}{2} - \frac{2}{3}\right) - \left(\frac{x_2^2}{2} - \frac{1}{6}\right) + \frac{7}{6} \\
&= \max(x_1, x_2) - \frac{x_1^2 + x_2^2}{2} + 2
\end{aligned}$$

In particular, if  $x_2 > x_1$ ,

$$\frac{\partial U_{12}(x_1, x_2)}{\partial x_1} = -x_1$$

so that  $U_{12}$  is not always increasing w.r.t. its two variables.

Hence the question of finding a decomposition into nondecreasing terms (which we call hereafter a *monotone* decomposition) is yet unsolved and as far as we know, its existence has not been studied. The next section tries to answer (at least partly) this question.

#### 4.4.3 Toward a monotone decomposition of a GAI model

We aim in this section at proposing some hints to define a decomposition of a GAI model in which all of its terms are nonnegative and nondecreasing.

The idea of ANOVA is valuable: to define decomposition in which  $F_i$  is the effect of moving  $x_i$  alone in  $F$ , and  $F_{i,j}$  is the effect of moving  $x_i$  and  $x_j$  simultaneously, getting rid of the effect of varying only one of these variables. Higher order terms are defined likewise. This condition of  $F_S$  representing the sole contribution of varying all variables in  $S$  simultaneously can be put as an orthogonality condition:

$$\mathbb{E}_{Z_i}[F_S | Z_{S \setminus i}] = 0 \quad \forall i \in S.$$

We wish to keep the previous idea of orthogonality but represented in another way. Orthogonality will be written in terms of the difference with a reference situation – depicted as alternative  $\mathbb{O}$  – as in Braziunas' approach. Following the ANOVA decomposition, a two-additive GAI model is written as

$$U(x) = u_\emptyset + \sum_{i \in N} u_i(x_i) + \sum_{\{i,j\} \subseteq N} u_{i,j}(x_i, x_j). \quad (31)$$

Term  $u_i$  should be the sole contribution of attribute  $x_i$ , removing the constant part. With a reference to option  $\mathbb{O}$ , we obtain the condition

$$u_i(\mathbb{O}_i) = 0. \quad (32)$$

Term  $u_{i,j}$  shall not depict the effect on  $U$  of the variation of only  $x_i$  or  $x_j$ . Hence

$$u_{i,j}(x_i, \mathbb{O}_j) = 0 \quad \forall x_i \in X_i \quad (33)$$

$$u_{i,j}(\mathbb{O}_i, x_j) = 0 \quad \forall x_j \in X_j \quad (34)$$

By (32), (33) and (34), the term  $u_\emptyset$  is the constant term:

$$u_\emptyset = U(\mathbb{O}). \quad (35)$$

In order to fulfill (32), (33), (34) and (35), nonnegativity and monotonicity of each  $u_S$  (i.e. (27), (28) and (29)), the  $\mathbb{O}$  alternative shall be the least element on each attribute. The next example shows that it is not always possible to find a decomposition fulfilling all previous conditions.

*Example 4 (Ex. 2 continued).* Let us consider (22):  $U(x_1, x_2) = x_1 + \max(x_1, x_2)$ . Let us try to find a decomposition of  $U$  in terms of  $u_\emptyset$ ,  $u_1(x_1) = x_1 + v_1(x_1)$ ,  $u_2(x_2) = v_2(x_2)$  and  $u_{1,2}(x_1, x_2) = \max(x_1, x_2) - v_1(x_1) - v_2(x_2)$  (where functions  $v_1$  and  $v_2$  are unknown) satisfying all previous conditions.

As  $u_2$  is monotone,  $v_2'(x_2) \geq 0$ . For  $x_1 > x_2$ ,  $\frac{\partial u_{1,2}}{\partial x_2}(x_1, x_2) = -v_2'(x_2) \geq 0$ . Hence  $v_2$  is constant. We obtain  $v_2 \equiv 0$  by (32).

Hence by (33),  $u_{1,2}(x_1, \mathbb{O}_2) = x_1 - v_1(x_1) = 0$ , so that  $v_1(x_1) = x_1$  and  $u_{1,2}(x_1, x_2) = \max(x_1, x_2) - x_1$ . For  $x_1 < x_2$ ,  $\frac{\partial u_{1,2}}{\partial x_2}(x_1, x_2) = -1$ , which violates the monotonicity of  $u_{1,2}$ .

We need to relax hard constraints (32), (33), (34) and (35), and transform them into an optimization objective. Finally, one aims at finding a decomposition of the form (31), where variables  $u_\emptyset$ ,  $u_i$ ,  $u_{i,j}$  are found by minimizing

$$|u_\emptyset|^p + \sum_{i \in N} |u_i(\mathbb{O}_i)|^p + \sum_{\{i,j\} \subseteq N} \left[ \sum_{x_i \in X_i} |u_{i,j}(x_i, \mathbb{O}_j)|^p + \sum_{x_j \in X_j} |u_{i,j}(\mathbb{O}_i, x_j)|^p \right] \quad (36)$$

where  $p > 0$  is a fixed power factor, under the constraints that (31) shall hold for every  $x \in X$ , the  $u_i$ 's and  $u_{i,j}$ 's are nonnegative and monotone (i.e. (27), (28) and (29)). The minimization of (36) is indeed a relaxation of a weak version of (32), (33), (34) and (35).

One can enrich (36) with other terms – for instance, the entropy of the unknowns not represented in (36). If the obtained functional is convex, we get a unique solution to the optimization problem.

## References

- F. Bacchus and A. Grove. Graphical models for preference and utility. In *Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 3–10, Montreal, Canada, July 1995.
- D. Bigot, H. Fargier, J. Mengin, and B. Zanuttini. Using and learning GAI-decompositions for representing ordinal rankings. In *Eur. Conf. on Artificial Intelligence (ECAI)*, Montpellier, France, August 2012.

- C. Boutilier, F. Bacchus, and R. Brafman. UCP-networks: A directed graphical representation of conditional utilities. In *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence (UAI-01)*, pages 56–64, Seattle, 2001.
- D. Bouyssou and M. Pirlot. Conjoint measurement tools for MCDM. In S. Greco, M. Ehrgott, and J. Figueira, editors, *Multiple Criteria Decision Analysis*, pages 97–151. Springer, 2016.
- D. Braziunas. *Decision-theoretic elicitation of generalized additive utilities*. PhD thesis, University of Toronto, 2012.
- G. Choquet. Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295, 1953.
- A. Fallah Tehrani, W. Cheng, K. Dembczynski, and E. Hüllermeier. Learning monotone nonlinear models using the Choquet integral. *Machine Learning*, 89:183–211, 2012.
- P. Fishburn. Interdependence and additivity in multivariate, unidimensional expected utility theory. *International Economic Review*, 8:335–342, 1967.
- R. A. Fisher and W. A. Mackenzie. The manurial response of different potato varieties. *J. of Agricultural Science*, 13:311–320, 1923.
- M. Grabisch.  $k$ -order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189, 1997.
- M. Grabisch and Ch. Labreuche. Capacities on lattices and  $k$ -ary capacities. In *3d Int. Conf. of the European Soc. for Fuzzy Logic and Technology (EUSFLAT 2003)*, pages 304–307, Zittau, Germany, September 2003.
- M. Grabisch and Ch. Labreuche. A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *4OR*, 6:1–44, 2008. doi 10.1007/s10288-007-0064-2.
- M. Grabisch and Ch. Labreuche. Monotone decomposition of 2-additive Generalized Additive Independence models. *Mathematical Social Sciences*, to appear.
- M. Grabisch, J.-L. Marichal, R. Mesiar, and E. Pap. *Aggregation functions*. Number 127 in Encyclopedia of Mathematics and its Applications. Cambridge University Press, 2009.
- M. Grabisch, Ch. Labreuche, and M. Ridaoui. On importance indices in multicriteria decision making. *working paper*, submitted.
- S. Greco, V. Mousseau, and R. Słowiński. Robust ordinal regression for value functions handling interacting criteria. *Eur. J. of Operational Research*, 239(3):711–730, 2014.
- C. R. Hsiao and T. E. S. Raghavan. Multichoice cooperative games. In B. Dutta, editor, *Proc. of the Int. Conf. on Game Theory and Economic Applications*, New Delhi, India, 1990.
- C. R. Hsiao and T. E. S. Raghavan. Shapley value for multichoice cooperative games, I. *Games and Economic Behavior*, 5:240–256, 1993.
- F. Klijn, M. Slikker, and J. M. Zarzuelo. Characterizations of a multichoice value. *Int. J. of Game Theory*, 28:521–532, 1999.
- Ch. Labreuche and M. Grabisch. Using multiple reference levels in multi-criteria decision aid: the Generalized-Additive Independence model and the Choquet integral approaches. *Eur. J. of Operational Research*, to appear.

- H. Peters and H. Zank. The egalitarian value for multichoice games. *Annals of Operations Research*, 137:399–409, 2005.
- M. Ridaoui, M. Grabisch, and Ch. Labreuche. Axiomatization of an importance index for generalized additive independence models. In *Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2017)*, Lecture Notes in Artificial Intelligence, pages 340–350, Lugano, Switzerland, July 2017a. Springer.
- M. Ridaoui, M. Grabisch, and Ch. Labreuche. An alternative view of importance indices for multichoice games. In *5th Int. Conf. on Algorithmic Decision Theory*, Luxembourg, October 2017b.
- M. Ridaoui, M. Grabisch, and Ch. Labreuche. An interaction index for multichoice games. arXiv:1803.07541, 2018.
- G. C. Rota. On the foundations of combinatorial theory I. Theory of Möbius functions. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 2:340–368, 1964.
- L. S. Shapley. A value for  $n$ -person games. In H. W. Kuhn and A. W. Tucker, editors, *Contributions to the Theory of Games, Vol. II*, number 28 in *Annals of Mathematics Studies*, pages 307–317. Princeton University Press, 1953.
- A. van den Nouweland, S. Tijs, J. Potters, and J. Zarzuelo. Cores and related solution concepts for multi-choice games. *ZOR – Mathematical Methods of Operations Research*, 41:289–311, 1995.
- J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, 2nd edition, 1947.