The Construction of a Vague Fuzzy Measure Through L^1 Parameter Optimization

Evgeni Dimitrov, Hayden Schaeffer, David Wen,
Sandra Rankovic, Kizza Nandyose, and Olivier Thonnard

Abstract—
This paper presents a method to construct an aggregation function, reflecting a complex set of initial user preferences, which can be used in the framework of multi-criteria decision making. We consider problems where the decision maker can provide information about the importance and interactions between criteria, as well as a desired portion of criteria to be satisfied. The proposed aggregation process is a vague Choquet integral, whose parameters are constructed in two steps. First, we solve a convex constrained L^1 optimization problem to obtain a fuzzy measure reflecting the importances and interactions between the criteria. Then the measure is transformed by a monotonic mapping to include vague information on what portion of criteria has to be satisfied. The proposed approach provides an automated construction of an aggregation function, which is completely free of data learning and manual processing. In addition, this method provides a novel fuzzy measure that integrates two different classes of information - importance/interactions of criteria and vague statements.

Index Terms—**Multiple criteria decision aid, Choquet integral, L1 optimization, vague statements, importance and interactions**

I. INTRODUCTION

MULTI-criteria decision analysis (MCDA) is a subdiscipline of operations research that focuses on making decisions based on the combined information provided by a set of criteria. It often involves choosing the best alternative from a given set of options by integrating several expert opinions. MCDA has found numerous applications in various areas such as transportation, sustainable energy, corporate and financial decision making (see e.g., [45], [9], [49], [50]) and recently cyber crime attack attribution problems [40]. A particular method of combining information in MCDA problems is through aggregation functions. Aggregation functions (or aggregation operators) can be applied in situations where one has to choose an optimal element from a given set of actions or objects, which can be compared with respect to different features. A typical problem involves having some numeric representation (*scores*) of how well each object satisfies each of a set of criteria. Once those numbers are known, one's aim is to combine the given scores, i.e. aggregate them into a global value, which provides an overall measure of how good the object is. The global scores can then be used to rate the different alternatives and consequently decide which is optimal.

The combination of scores is achieved by an aggregation function, which in all generality is simply a multi-variable

function, that has a real output. Some typical examples of aggregation functions are the arithmetic and weighted mean. Those are usually used in problems where one can rate the importance of the different criteria using *weights*, which reflect how much the final aggregated value depends on each individual score (see e.g. [34], [44]). When applying linear models like the weighted mean one assumes that different features are independent, however in more complex environments they could interact (see e.g. [30], [11], [39]). Specifically, it could be the case that several criteria point to similar information regarding the object and while each may be individually very important, their combined effect on the aggregated value should not be much higher than the effect of each independently. In the case described we would call the criteria or features *redundant*. Conversely, one could have that several criteria capture complementary information so that their overall effect is much higher than the individual effect of each. We refer to the latter as being *complementary* or *synergetic*.

The typical example that appears in the literature to illustrate such interactions is the student ranking problem (see [16]). In this problem one has all the grades in various subjects of each student and wishes to rank the latter based on their overall performance. The additional assumption made is that subjects are not independent. For example, if a student is good at mathematics he/she is very likely to be good at physics as well. Thus the two grades in mathematics and physics point to similar student qualities and are redundant. Conversely, good grades in mathematics and literature indicate that a student is "well rounded" and are complementary. Linear models like the weighted average are incapable of capturing the complex relationships we just described, which necessitates the application of different aggregation functions. The typical aggregation function used for reflecting criteria interactions is the *Choquet integral* (see [16], [30], [32], [21]). The formulation of the Choquet integral makes it particularly useful for modeling complex criteria relationships and it has become a central tool for addressing problems in MCDA, finding numerous practical applications (see the surveys in [32], [21], [26]).

Another variation of the aggregation problem involves the case when one rates objects based on the portion of criteria they satisfy without any preference on which exactly. For example, a decision maker (DM) might value objects, which satisfy at least half of a set of criteria well, regardless of how they perform on the other half. Alternatively, a DM might wish to disregard the best and/or most poorly satisfied

criterion when evaluating an object due to some possible bias. In the examples described, the rating of an element only depends on the set of scores it has for each criterion and is invariant under permutation of those scores. This introduction of nonlinearity renders the weighted average inapplicable and several operators have been proposed for this kind of problem such as the ordered weighted average (OWA) and the weighted OWA (WOWA) [46], [43]. Operators in the OWA family have been used for rating objects based on preferences of the form: "assign a high aggregated value if at least half of the scores are high". The example just given is often referred to as a *vague statement* and one can obtain different vague statements by replacing "at least half", with "most", "some", "all" etc. *Linguistic quantifiers* such as "most" and "some" are often open to interpretation which is why statements specified by them are called vague [48]. Over the last couple of decades OWA operators have found numerous applications in various fields like neural networks ([7], [47]), geographical information systems ([24], [35]) and group decision making under linguistic assessments ([22], [23]). We remark that both the OWA and WOWA are special cases of the Choquet integral, however the two types of aggregation operators have largely been used in different kinds of problems (see [16], [21]).

The two variations of the aggregation problem presented above have been extensively studied and analyzed in [48], [12], [14], [18], [33], [16], [21]; however, there is no one universally accepted method for approaching problems that present both types of preferences. It is thus the purpose of this paper to develop an aggregation function that is capable of reflecting a complex set of user preferences, which include:

- relative importance of different criteria
- interactions between criteria
- vague statements.

In particular we will assume that the user (decision maker) can rate the importance of different criteria, has information about the way they interact and has the additional preference of a vague statement. In our proposed approach, which will be described in more detail in the next section, we will try to combine some of the benefits of the OWA and the Choquet integral. We will take advantage of the ability of the OWA and the Choquet integral to model vague statements and interactions between criteria respectively and incorporate those features into a final aggregation function that reflects all of the above user preferences.

In order to take advantage of the two operators we first need to construct the sets of parameters that define them. In many cases the parameters are designed manually by an expert of the decision problem, which is being modeled (for some recent examples see [41], [40]). However, if the number of criteria is N we have that the OWA is defined by N weights, while the Choquet integral by 2^N parameters. The high parameter complexity of the models makes manual construction extremely difficult as N becomes large. Thus for practical purposes we design our method of construction to be *automated*.

The framework, which we will use for designing the parameters of the Choquet integral is *optimization*. Some applications of optimization for finding the parameters of the

Choquet integral can be found, among many others, in [14], [33], [20], [29], [32], [25] . In most of these papers the authors assume that one has a set of rated examples, which can be used as a basis for the construction of a fuzzy measure. In practice this may not be true as can be seen in [41], [40], where due to insufficient data the authors constructed their parameters manually. We thus develop our model to be free of any data learning. Avoiding data learning makes our approach much more applicable to areas where there is a lack of sufficient data or where data quickly becomes outdated and unusable for any decision making. We only assume that the user can provide information on the preferences listed above and not necessarily a rating of different examples, although this modification can easily be incorporated.

Optimization (also known as energy minimization) is used in many fields to solve ill-posed inverse problems. Typically, one wants to solve the problem $Au = f$, where f is given but A is ill-conditioned. The problem is solved by minimizing the following quantity:

$$F(u) = ||u|| + \frac{\lambda}{2}||Au - f||_2^2$$

where $|| - ||$ is a regularizer (commonly in the form of a norm or semi-norm). In the fields of compressive sensing (see [10], [28], [4]) and image processing (see [27], [36], [3]), sparse regularization is frequently used in order to reconstruct a vector u, which has few non-zero elements. The optimal regularizer is the L^0-"norm"; however, a convexified version can be found by replacing it with the L^1-norm. The resulting minimization is as follows.

$$\begin{aligned} \min_u F(u) &= \min_u ||u||_1 + \frac{\lambda}{2}||Au - f||_2^2 \\ &= \min_u \sum_j |u_j| + \frac{\lambda}{2}||Au - f||_2^2 \end{aligned}$$

Other L^1 type regularizers include the total variation (TV) semi-norm (*i.e.* $||\nabla u||$) and the nuclear norm (*i.e.* the sum of the singular values). The TV semi-norm induces sparsity in the jump set of the data (see [36]) while the nuclear norm induces pattern sparsity in large data sets (see [5], [37]).

In this paper, we will use the L^1 norm to regularize the interaction parameters within the model of the Choquet integral, which is supported by the theoretical well-posedness of a 2-additive fuzzy measure (see Section III). The parameter f is provided by the user and represents his/her belief on what the interactions should be, while the matrix A is a binary operation, which is 1 if an interaction is given and 0 otherwise.

The remainder of the paper is organized as follows. In Section II we describe the format of the preferences provided by the user as well as the general strategy of constructing our final aggregation function. In the third section we summarize some of the basic facts and definitions pertaining to the Choquet integral. Section IV develops the optimization framework we employ to construct the fuzzy measure, reflecting criteria importance and interactions. In the following section we design the weights for the OWA, based on the vague statement. Subsequently, we combine the two sets of parameters into a vague fuzzy measure, which models our final aggregation

function. In Section VI we analyze the performance of our operator on two examples and conclude with some closing remarks in Section VII.

II. OUTLINE OF CONSTRUCTION METHOD

As we mentioned in the previous section our goal is to construct an aggregation function that captures a complex set of user preferences including relative importance and interactions of criteria and vague statements. We will assume that a user presents this information in the following form:

- **Importance**: The relative importance of criteria is provided via statements like "criterion 1 is twice as important as criterion 2".
- **Interactions**: The interaction between two criteria i and j is given via a number between -1 and 1. Negative values are interpreted as i and j being redundant, i.e. pointing to similar information, while positive values are interpreted as the two criteria being synergetic, i.e. pointing to complementary information. The value 0 implies that the two criteria are independent.
- **Vague statement**: A linguistic quantifier is submitted and it determines how many scores need to be high in order to maximize the final aggregated value.

As mentioned in the introduction, the formulation of the Choquet integral makes it exceptionally suitable for modeling the importance and interactions of criteria, while the OWA appropriately models vague statements. Thus as an initial step we build the parameters underlying these two aggregation functions. The fuzzy measure for the Choquet integral is constructed based on the information given for the importance and interactions, while the weights for the OWA are built based solely on the vague statement.

For this first step we will assume a 2-additive model for the Choquet integral, which will provide us with a sufficiently powerful framework for representing the types of preferences described above. The 2-additive Choquet model is specified by two kinds of parameters called *importance indices* and *interaction indices*. As the names suggest they measure the importance of each criterion and the interaction between pairs of criteria respectively. We will construct these parameters, by minimizing an energy function that reflects properties that we view as desirable for an optimal solution. Subsequently we design weights for an OWA operator. We will use *RIM quantifiers* of the truncated Gaussian distribution to construct those weights, which provides a natural and well behaved solution.

In the second step of construction we use both sets of parameters designed in the first step to create a fuzzy measure, which will be a basis for the Choquet integral that is our final aggregation function. For clarity, we will refer to the latter as a *vague* fuzzy measure and a *vague* Choquet integral to stress their dependence on the vague statement. The purpose of the vague Choquet integral is to combine the information reflected by the parameters of the OWA and the Choquet integral. It acts by increasing the aggregated value of examples which both models render good and penalizing the aggregated score of examples on which the two disagree.

The next section provides some of the basic definitions and results for the Choquet integral and thus may be skipped by those who are familiar with the concept. We refer the reader to [1], where one can find additional properties of the discrete Choquet integral as well as the OWA function.

III. PRELIMINARY DISCUSSION ON FUZZY MEASURES

For the rest of the paper, we assume $\mathcal{N}$ to be a set of n elements, $2^{\mathcal{N}}$ to be the power set of $\mathcal{N}$ and z to be a vector in $\mathbb{R}$, whose entries are real values between 0 and 1.

Definition 1. *A* capacity [8] *or* fuzzy measure [38] *is a set function $\nu : 2^{\mathcal{N}} \to [0,1]$, satisfying the following properties:*
1) $\nu(\emptyset) = 0$
2) $\mathcal{A} \subseteq \mathcal{B}$ implies $\nu(\mathcal{A}) \leq \nu(\mathcal{B})$
The fuzzy measure is normalized if in addition $\nu(\mathcal{N}) = 1$.

Definition 2. *The (discrete) Choquet integral of an input vector* **z** *with respect to a fuzzy measure ν is given by*

$$C_\nu(\mathbf{z}) = \sum_{i=1}^{n} z_{(i)}[\nu(\{j|z_j \geq z_{(i)}\}) - \nu(\{j|z_j \geq z_{(i+1)}\})] \quad (1)$$

Where $z_{(1)} \leq z_{(2)} \leq ... \leq z_{(n)}$, i.e. $z_{(i)}$ is the ith largest component of the input vector **z**. *In addition, we use the convention that $z_{(n+1)} = \infty$.*

Within this paper fuzzy measures are always assumed to be normalized. We in addition adopt the notation $\mathbf{z} \searrow$ to mean the vector, whose components are the entries of **z** sorted in descending order.

The formulation of the Choquet integral, requires the definition of $2^n - 2$ parameters, corresponding to the subsets of $\mathcal{N}$ (excluding $\emptyset$ and $\mathcal{N}$, whose values are fixed to be 0 and 1 respectively). The exponential complexity of the model often makes it difficult to construct parameters that exhibit certain desirable properties, which necessitates the use of simpler formulations that are easier to mold. A particular simplification introduced in [18] is the notion of *k*-additive fuzzy measures. *k*-additive fuzzy measures are families of fuzzy measures, ranging from the additive case ($k = 1$) to the general case ($k = n$). The formal definition is as follows.

Definition 3. *A fuzzy measure ν is said to be at most k-additive ($1 \leq k \leq n$) if its Möbius transform satisfies*

$$\mathcal{M}_\nu(\mathcal{A}) = 0$$

for any subset $\mathcal{A}$ with more than k elements, $|\mathcal{A}| > k$. A fuzzy measure ν is k-additive if in addition there exists a subset $\mathcal{B} \subseteq \mathcal{N}$ with k elements such that $\mathcal{M}_\nu(\mathcal{B}) \neq 0$, where the Möbius transform *of a fuzzy measure ν is a set function defined for every $\mathcal{A} \subseteq \mathcal{N}$ as*

$$\mathcal{M}_\nu(\mathcal{A}) = \sum_{\mathcal{B} \subseteq \mathcal{A}} (-1)^{|\mathcal{A} \setminus \mathcal{B}|} \nu(\mathcal{B})$$

In this paper we are primarily focusing on 2-additive fuzzy measures, which extend the basic additive model by allowing for interactions between elements of $\mathcal{N}$. In the case of a 2-additive measure one can obtain a different representation of

the Choquet integral, which involves an easier to interpret set of parameters called the *importance* and *interaction indices*.

Definition 4. *The importance (Shapley) index of an element i $\in \mathcal{N}$ w.r.t. a fuzzy measure ν is given by*

$$I_i = \sum_{\mathcal{A} \subseteq \mathcal{N} \setminus i} \frac{(n - |\mathcal{A}| - 1)! |\mathcal{A}|!}{n!} [\nu(\mathcal{A} \cup \{i\}) - \nu(\mathcal{A})]$$

The Shapley value *is the vector* $I_s = (I_1, ..., I_n)$.

Definition 5. *The interaction index between two elements i,j $\in \mathcal{N}$ w.r.t. a fuzzy measure ν is given by*

$$I_{ij} = \sum_{\mathcal{A} \subseteq \mathcal{N} \setminus \{i,j\}} \frac{(n - |\mathcal{A}| - 2)! |\mathcal{A}|!}{(n-1)!} [\nu(\mathcal{A} \cup \{i,j\}) -$$
$$\nu(\mathcal{A} \cup \{i\}) - \nu(\mathcal{A} \cup \{j\}) + \nu(\mathcal{A})]$$

An alternative representation of the interaction index between i and j can be given in terms of the Möbius transform of the fuzzy measure ν:

$$I_{ij} = \sum_{\mathcal{B} | i,j \in \mathcal{B}} \frac{1}{|\mathcal{B}| + 1} \mathcal{M}_\nu(\mathcal{B})$$

An importance index measures the contribution of a specific element in all possible coalitions, i.e. the "benefit" of adding that element to an already constructed subset of $\mathcal{N}$. Similarly, the interaction index measures the average contribution of adding a pair of elements to a given set, as opposed to adding just one of the elements. If $I_{ij} > 0$ the elements i and j are said to be *synergetic* and if $I_{ij} < 0$ they are said to be *redundant*. The notion of the interaction index has been extended by Grabisch in [17], to represent interaction among arbitrary subsets of $\mathcal{N}$. However, since we are working with a 2-additive fuzzy measure, we only concern ourselves with the sets of parameters - I_i and I_{ij}. In fact, it can be proved (see for example [18], [21]) that these parameters completely determine the fuzzy measure if it is 2-additive and one can represent the Choquet integral alternatively in the following form.

$$C_\nu(\mathbf{z}) = \sum_{i \in \mathcal{N}} z_i I_i - \frac{1}{2} \sum_{\{i,j\} \subset \mathcal{N}} I_{ij} |z_i - z_j| \qquad (2)$$

We end this section with a set of useful identities when ν is 2-additive, which are going to be used extensively further in the paper. The properties listed below are easily verified from the given definitions.

$$\sum_{i=1}^{n} I_i = 1 \iff \nu \text{ is normalized}$$
$$\qquad (3)$$
$$I_i \geq \frac{1}{2} \sum_{i \neq j} |I_{ij}| \quad \forall i \in \mathcal{N} \iff \nu \text{ is monotone}$$

Notice that the condition that the fuzzy measure is monotone is related to the L^1 norm of the interactions. This relationship provides support that L^1 regularization is natural for optimizing the parameters of a Choquet integral.

IV. OPTIMIZATION AND FUZZY MEASURES

As mentioned in the first two sections it is our desire to build an aggregation function for modeling a wide range of preferences of a DM. Since we are working with a rather complex set of preferences, we split this problem in two parts. In the first we construct a 2-additive fuzzy measure, which reflects the relative importance and interactions among criteria as well as a set of weights that capture the provided vague statement. In the second step we use both sets of parameters and based on them construct a *vague* fuzzy measure that captures all user preferences. This section describes how we obtain the 2-additive fuzzy measure as the unique minimizer of a convex optimization problem.

The application of optimization for designing fuzzy measures has been studied extensively in previous works on MCDA [33], [14], [20], [29], [32], [31], [25]. Due to the complex nature of the Choquet integral one often needs an automated way to construct its parameters. A common approach is to define an objective function, reflecting some desired property. The function is then optimized under a set of constraints, imposed by the preferences of the DM and by the theoretical constraints of the model. In [33],[14], [20] the Choquet integral was defined as the unique minimizer of the total squared error, over a set of alternatives. In particular, the DM assigns some values to different options and then the Choquet integral is determined, by the fuzzy measure, which gives the smallest deviation from those values. A different approach, adopted in [29], [32], [31], maximizes the difference in overall scores among alternatives. I.e. given some ordinal information for a set of options, one finds the measure that maximizes the overall distance between these options, thus optimally differentiating between them.

The methods described above depend strongly on learning data provided by the DM. Since we desire our construction to be independent of data learning, we cannot use either of them and thus approach the problem differently. Our aim is to only use information provided by the DM on the relative importance and interactions among criteria, without asking him/her to rate alternatives in any way. The objective function that we optimize is then going to model the general structure of the measure, subject to those constraints. In this context, our approach is similar to the one used by Kojadonovic in [25] with two major differences. Firstly, the energy function that we develop aims to obtain a sparse representation of the inter-actions among criteria, for reasons that will be explained later in this section. Secondly, we strongly take advantage of the 2-additive form of the measure, and do not aim to maximize its uncertainty as in [25], since not all available information has been introduced during the first step of construction. Given our distinct goals, we propose a novel objective function and formulation of the problem.

A. The Data

We assume that the DM provides the following set of preferences, based on his/her *domain knowledge*:

- Linear relationships between the importances of criteria

- Values for the interactions between two criteria on the scale $[-1, 1]$.

The importances of criteria are connected via statements like "Criterion 1 is twice as important as criterion 2". The numerical value entered for the interaction between criteria i and j is denoted by I_{ij}^0. If no information is provided on how i and j interact the value I_{ij}^0 is set to zero (i.e. the criteria are independent). A positive value implies that the DM believes the criteria are synergetic and a negative value - that they are redundant. We interpret the provided cardinal values as the DM's "initial guess" on the interaction indices in our 2-additive fuzzy measure.

It is difficult to motivate how the DM comes up with the above values for the interactions. This problem also appears in [25] where the author includes equations such as $I_{ij} = m$ (where m is some numerical constant) within the set of possible constraints. These equations uniquely determine the interaction indices, although the author does not restrict himself to 2-additive fuzzy measures. In our case, we aim to obtain a 2-additive fuzzy measure, which imposes much stricter conditions than the general case. In order to overcome this issue, we relax the condition of setting the interaction indices exactly as prescribed by the user. Instead, we add a fidelity term in our energy function that minimizes the distance to the desired user values, while maintaining the 2-additive structure of the solution. The fidelity term ensures that the values obtained for the interaction indices are close to the one provided by the user, without necessarily imposing strict equality. The motivation for this relaxation of the constraints lies in our assumption that the initial parameter set does not necessarily define a well-posed fuzzy measure.

B. Feasible Set

The next step in our problem is deriving the constraints in our optimization problem from the preferences of the DM. Similarly to [30] and [25], we translate preferences on criterion importance into linear constraints using the importance indices of a fuzzy measure. For example $I_1 - 2I_2 = 0$, would be the interpretation of "criterion 1 is twice as important as criterion 2". In general, the k-th preference can be written as $\sum_{i=1}^{n} a_{ik} I_i = 0$, resulting in a linear system with $A = [a_{ik}]$ as the linear operator. The parameters I_{ij}^0 will be used in the definition of our objective function, and not used as constraints.

Remark: Our approach does not require that the relationship between the importance indices is given via linear equalities. The approach admits any dependency that leads to a convex set of feasible values. In particular, one can define relationships via inequalities of the form $I_1 - I_2 \geq 0$, to indicate that the first criterion is more important than the second or as in [30] and [25] use $I_1 - I_2 \geq \delta$, with some predefined small positive value δ.

From the theory, the constructed measure needs to be normalized and monotone in order to be well posed. The latter conditions were provided in equation 3 at the end of the previous section. Altogether, the (convex) set of feasible

solutions (I_i, I_{ij}) is:

$$\mathcal{A} = \left\{ \sum_{i=1}^{n} a_{ik} I_i = 0 \ \forall k, \ \sum_{i=1}^{n} I_i = 1, \ I_i \geq \frac{1}{2} \sum_{i \neq j} |I_{ij}| \right\} \quad (4)$$

In general, as long as the $Ker(A)$ is "large enough," the set $\mathcal{A}$ stays non-empty. In practice, as long as the equality relations between the importances indices remain consistent, a solution will exist.

C. Objective function for parameter estimation

We propose the following objective (energy) function to be minimized.

$$F(\{I_i\}_{i \in \mathcal{N}}, \{I_{ij}\}_{i,j \in \mathcal{N}}) = \alpha \sum_{i \in \mathcal{N}} I_i \log I_i + \\ \sum_{i,j \in \mathcal{N}} \left(|I_{ij}| + \beta (I_{ij} - I_{ij}^0)^2 \right), \quad (5)$$

where α and β are fixed positive constants. Both the function and feasible set is convex, providing a unique minimizer (when one exists). Also note that although the energy function does not couple I_i and I_{ij}, the constraints that they lie in $\mathcal{A}$ does.

The first term of equation 5, represents the entropy of the importance indices, which is chosen to provide an appropriate distribution for the values of I_i even when little information is provided. This term also discourages I_i from being zero, therefore every criterion is taken into account in the decision process. The entropy term ensures all criteria are considered nearly-equal in terms of importance when no information is provided to the contrary.

The second term in equation 5 is the L^1-norm of the interaction parameters, encouraging sparsity in the interactions I_{ij}. A sparse distribution of interaction parameters is well motivated by both theory and practice. In terms of theory, being a well-posed fuzzy measure translates to an L^1 constraint on I_{ij} (in terms of I_i). In addition, the nature of the 2-additive fuzzy measure implies that complex interactions (higher order behaviors) between the criteria must be ignored for well-posedness. In terms of practice, one assumes that only some criteria interact, otherwise the given criteria may not be reasonable. A low amount of interacting criteria implies sparsity of the interaction parameters. In general, the interactions of higher (relative) magnitude are more influential in the overall aggregation process, so that the idea of sparsity is consistent. Thus having an L^1-norm in the energy function allows us to obtain a solution that captures the basic relationships between criteria, without leaving the framework of 2-additivity.

As mentioned before the parameters I_{ij}^0 represent the user's belief of what the interactions should be. Of course, it could be the case that the input values are not feasible for a 2-additive fuzzy measure or inconsistent in some other way. In this regards, we include an L^2 fidelity term to keep our solution close to, but not exactly equal to, the initial data set. This norm is used in L^2 data fitting, for example, least squares fitting. Minimizing this norm maintains a level of closeness between the input and output interactions. In general, any p-norm with $p > 1$ can be used; however, L^2 is chosen since

in the unconstrained case, the minimizers are found by soft thresholding, and this provides the desired sparsity behavior.

Altogether, the proposed optimization is as follows.

$$\underset{\{I_i\}_{i\in\mathcal{N}},\{I_{ij}\}_{i,j\in\mathcal{N}}}{\text{minimize}} F(\{I_i\}_{i\in\mathcal{N}},\{I_{ij}\}_{i,j\in\mathcal{N}})$$

$$\text{subject to}\ \ (\{I_i\}_{i\in\mathcal{N}},\{I_{ij}\}_{i,j\in\mathcal{N}})\in\mathcal{A}$$

The above is a convex optimization problem, which has a unique solution and can be solved using well-developed methods, for example, Lagrange multipliers or Bregman iterations. In order to provide some intuition on the behavior of the solution we consider the case when the set of linear constraints uniquely determine the importance indices. In that case the energy function reduces to

$$\sum_{i,j\in\mathcal{N}}\left(|I_{ij}|+\beta(I_{ij}-I_{ij}^0)^2\right)$$

The above function has a well-known minimizer in the unconstrained case given by

$$I_{ij}=shrink(I_{ij}^0,1/2\beta),$$

where the function $shrink(x,\lambda)$ for $\lambda>0$ is given by

$$shrink(x,\lambda)=\left\{\begin{array}{ll} x-\lambda & \text{if } x>\lambda \\ x+\lambda & \text{if } x<-\lambda \\ 0 & \text{otherwise.} \end{array}\right.$$

Using the above one can easily verify that in our optimization problem if $1/2\beta\geq|I_{ij}|^0$, then the unique minimizer necessarily satisfies $I_{ij}=0$. The latter indicates how in general a sparse interaction profile is obtained. Essentially, all interactions of small magnitude are automatically set to 0. The $shrink$ function plays an important role in L^1 regularization problems especially in methods such as the Split Bregman [13].

The exact behavior of the solution of our optimization problem is not completely clear, although some general tendencies such as the one above describe it to some extent. In Section VI we look into two specific examples of the optimization problem, which provide additional intuition about the minimizer. In both examples we will apply our model with $\alpha=1$ and $\beta=5$. Based on the above remarks, $\beta=5$ implies that all interactions for which $|I_{ij}^0|<0.1$ will automatically be set to zero, as we consider them too small. The full understanding of the effect of the parameter α requires further research. In general, higher values give precedence to well-spread importance indices, while smaller values impose closer adherence to the values given by the user.

V. A Vague Fuzzy Measure

Having built our fuzzy measure reflecting criteria importance and interactions we now proceed with modeling the vague statement, given by the DM. The typical way vague statements have been modeled in the literature on MCDA is by using the ordered weighted average (OWA) and the weighted OWA (WOWA) [48], [43], [40]. There exist well developed methods for constructing weighing vectors for the OWA and we choose that of RIM Quantifiers, based on the truncated Gaussian distribution [1].

The next step in our construction involves combining the two sets of parameters into a vague fuzzy measure, which will be the focus of this section. As mentioned earlier there is no accepted way of integrating these parameters and we thus develop a novel method for accomplishing this task. Using the definition of the WOWA as motivation we propose the following formulation.

Definition 6. *Given a fuzzy measure ν and a weighing vector $\mathbf{w}$ we define a fuzzy measure μ as*

$$\mu(\mathcal{A})=W(\nu(\mathcal{A})),\ \ for\ all\ \ \mathcal{A}\in 2^{\mathcal{N}}$$

where as usual $2^{\mathcal{N}}$ is the power set of $\mathcal{N}$, and W is a monotone non-decreasing function that interpolates the points $(\frac{i}{n},\sum_{j\leq i}w_j)$ together with the point $(0,0)$. Moreover, W is required to have the following two properties:

1) $W(\frac{i}{n})=\sum_{j\leq i}w_j,\ i=0,\cdots,n;$
2) *W is linear if the points $(\frac{i}{n},\sum_{j\leq i}w_j)$ lie on a straight line.*

With the above specific composition, we propose a vague fuzzy measure. Let us note that the obtained measure μ is indeed a fuzzy measure, since W is monotone and preserves zero. One can define many different functions W with the properties in the above definition such as a linear spline, a monotone quadratic spline, monotone cubic spline, etc. In particular, throughout this paper W is piecewise linear, interpolating the points $(\frac{i}{n},\sum_{j\leq i}w_j)$.

A. Constructing the weights

There are many choices for the weighing vector $\mathbf{w}$. Some typical examples seen in literature define $\mathbf{w}$ as the unique solution to a different optimization problem such as the *maximal entropy OWA* and the *minimal variance OWA* [12]. Yet other ways involve the use of *weight generating functions* [1]. In Definition 6 there is no restriction on what the weighing vector can be, so long as it is normalized and has nonnegative entries. Based on experiments, we favor the use of weight generating function, specifically, those based on the truncated Gaussian distribution.

Given a vague statement of the form "k out of n criteria are satisfied", we construct a Gaussian distribution with mean $m=\frac{2k-1}{2n}$ and variance σ. The weights are explicitly given as:

Definition 7. *If $\mathbf{w}$ is a weighing vector of size n, we have that*

$$w_i=\frac{1}{K}\int_{\frac{i-1}{n}}^{\frac{i}{n}}e^{\frac{-(x-m)^2}{2\sigma^2}}dx$$

where K is a normalization constant.

We choose $\frac{2k-1}{2n}$, because it lies precisely between $\frac{k-1}{n}$ and $\frac{k}{n}$, which would mean that the largest entry of $\mathbf{w}$ is precisely the k-th entry. This would imply that the ordered weighted average of a vector is high, precisely when it has at least k high entries. And we motivate the latter with a simple example.
Example: We are given the vague statement "at least 3 (out of 5) entries of $\mathbf{z}$ are high". From the above definition $m=0.5$ and for $\sigma=0.1$ and $\mathbf{w}=(0.001,0.157,0.683,0.157,0.001)$.

Then $\mathbf{w}$ has a clear emphasis on the third weight, which implies that the only way that $OWA_{\mathbf{w}}(\mathbf{z}) = <\mathbf{w}, \mathbf{z} \searrow>$ is high is if the third highest entry of $\mathbf{z}$ is high, meaning that $\mathbf{z}$ contains at least three high entries.

B. Properties of the vague Choquet integral

Based on the above formulation of the weighing vector $\mathbf{w}$ and the measure ν, constructed in the previous section, we can now construct μ (as in the formulation of Definition 6). Our final aggregation function is just a Choquet integral with respect to μ, which we denote by $C_{\nu,\mathbf{w}}$ to show its relationship to both $\mathbf{w}$ and ν.

We next wish to examine some of the basic properties of the newly constructed aggregation function. And we will begin by first looking into a couple of examples that compare how the function behaves, compared to $OWA_{\mathbf{w}}$ and C_{ν} (the Choquet integral with respect to ν). For the purposes of the following examples we will not use the optimization method of construction for ν, but rather just a simple handmade measure.

Example 2a: Suppose that we have three criteria - $\{K_1, K_2, K_3\}$ and we are given that $K_1 \sim K_2 \sim 2K_3$. In addition, let K_1 and K_2 have 40% synergy and let the vague statement provided be "at least two scores are high." We consider two vectors $\mathbf{x}$ and $\mathbf{y}$, such that $\mathbf{x} = [0.85\ 0.8\ 0.1]$ and $\mathbf{y} = [0.3\ 0.2\ 0.8]$.

The importance and interaction indices that define ν uniquely are: $I_1 = I_2 = 0.4, I_3 = 0.2$, $I_{12} = 0.4$ and $I_{23} = I_{13} = 0$. The weighing vector $\mathbf{w}$ in OWA is $\mathbf{w} = [0.1\ 0.8\ 0.1]$. We calculate the aggregated value of the two vectors using the different aggregation functions. The results are summarized in Table I.

Vector	$OWA_{\mathbf{w}}$	C_{ν}	$C_{\mathbf{w},\nu}$
$\mathbf{x}$	0.735	0.670	0.761
$\mathbf{y}$	0.340	0.340	0.256

TABLE I

As can be seen from Table I the aggregated value for $\mathbf{x}$ using $C_{\nu,\mathbf{w}}$ is higher than those obtained from using $OWA_{\mathbf{w}}$ and C_{ν}, while the one for $\mathbf{y}$ is lower. Both the OWA and the Choquet integral agree that $\mathbf{x}$ should receive a high aggregated score, which implies that $\mathbf{x}$ satisfies the constraints in both operators well. Looking at $\mathbf{x}$ we can clearly see that it has two high entries, which correspond to the two most important criteria (K_1 and K_2), which are also synergetic. In addition, $\mathbf{x}$ satisfies well the vague statement. Thus if we combine the two informations, we should obtain that $\mathbf{x}$ receives an overall high score, which is precisely the result we obtained. Similarly, for $\mathbf{y}$ we can see that its lowest values correspond to the most important criteria K_1 and K_2. In addition, $\mathbf{y}$ does not satisfy the given vague statement well, and thus should receive a low score as is the case. If a vector performs well on both kinds of information - importance/interactions and vague statement, then $C_{\nu,\mathbf{w}}$ assigns it a higher value than either $OWA_{\mathbf{w}}$ or C_{ν}. And if the vector performs badly on all counts it receives an even lower score. This is the desired behavior that we aim

Vector	$OWA_{\mathbf{w}}$	C_{ν}	$C_{\mathbf{w},\nu}$
$\mathbf{x}$	[0.1 0.8 0.1]	[0.2 0.6 0.2]	[0.06 0.88 0.06]
$\mathbf{y}$	[0.8 0.1 0.1]	[0.2 0.6 0.2]	[0.74 0.20 0.06]

TABLE II

for. Yet it is interesting to look into exactly how the vague fuzzy measure redistributes the weights in OWA, according to the information contained in the fuzzy measure ν. In order to answer this question let us look at the previous example in more detail.

Example 2b: As before $\mathbf{x} = [0.85\ 0.8\ 0.1]$ and $\mathbf{y} = [0.3\ 0.2\ 0.8]$. Let us consider the "weights" that each entry receives using the OWA, the Choquet integral and the vague Choquet integral. Specifically, for the two integrals we shall call the k-th "weight" the parameter that appears in front of the k-th index (i.e. $\nu(\{j|z_j \geq z_{(k)}\}) - \nu(\{j|z_j \geq z_{(k+1)}\})$ in the language of Definition 2). The weights for $\mathbf{x}$ and $\mathbf{y}$ are summarized in Table II.

Looking at Table II we can at least intuitively see how the proposed model operates. Consider the vector $\mathbf{x}$. We see that both $OWA_{\mathbf{w}}$ and C_{ν} agree that the second entry should have a high weight, which results in a very high weight in $C_{\mathbf{w},\nu}$. Similarly as the first and last entry of $\mathbf{x}$ receive low weights in both operators, those become even lower using the vague Choquet integral. Thus when both the OWA and Choquet integral agree that a certain entry should receive a high weight, this translates to an even higher weight in the vague Choquet integral and the converse is also true.

It is worth examining how the function behaves when the information from the OWA and the Choquet integral is not consistent. As an example consider the vector $\mathbf{y}$ where the aggregation functions disagree on the first two entries. In this case, the first weight in $C_{\mathbf{w},\nu}$ can be understood as a skewed down version of the weight in $OWA_{\mathbf{w}}$. The reason for the reduction is the low weight assigned by C_{ν} for the same entry. Conversely the second entry of $\mathbf{y}$ receives a higher weight in $C_{\mathbf{w},\nu}$ than in $OWA_{\mathbf{w}}$ because its weight in C_{ν} is high. In general weights in the OWA are skewed according to those in the Choquet integral, although the exact degree depends on the parameters of the models.

C. Some Important Cases

The proposed model has some additional nice properties, apart from the ones already described. In particular it reduces to some interesting aggregation functions in specific cases. We summarize these properties below.

1) If all interactions among criteria are zero, the vague Choquet integral becomes the $WOWA_{\mathbf{w},\mathbf{I}}$, where $\mathbf{w}$ is the weighing vector of OWA and $\mathbf{I}$ is the Shapley value.
2) If all interactions among criteria are zero and all criteria are equally important, the vague Choquet integral becomes the $OWA_{\mathbf{w}}$.
3) If $\mathbf{w} = [\frac{1}{n}, ..., \frac{1}{n}]$, the vague Choquet integral becomes a Choquet integral with respect to the fuzzy measure ν.

4) If $\mathbf{w} = [\frac{1}{n}, ..., \frac{1}{n}]$ and all interactions among criteria are zero, the vague Choquet integral becomes a weighted mean with a weighing vector $\mathbf{I}$.

5) If $\mathbf{w} = [\frac{1}{n}, ..., \frac{1}{n}]$, all interactions among criteria are zero and all criteria are equally important, the vague Choquet integral becomes the arithmetic mean.

All of the above properties can directly be verified by the definition of the vague fuzzy measure.

Remark: The two steps that we implemented in constructing our vague measure are completely autonomous, meaning that each step can be applied individually to different problems, for which they are suited.

VI. MODEL EVALUATION

In this section we present a practical implementation of our proposed approach. We consider several examples of synthetic data and describe the actions of the developed aggregation functions. Our goal is to illustrate the contribution of each step in our construction of the vague Choquet integral.

The proposed approach was implemented within MATLAB [51]. The optimization problem is solved using cvx - software package for MATLAB (for strictly convex optimization) [2].

A. Example I

The general setup is as follows. A DM chooses from a set of six options, which can be compared with respect to five criteria. To each of the six alternatives we associate a vector in $\mathbb{R}^5$, whose entries represent how well the element satisfies each criterion. As before we refer to the entries of the vectors as scores and all scores are assumed to be on the same scale $[0, 1]$. The DM provides a set of preferences and the task is to create an aggregation function modeling those preferences.

We denote the five criteria using the same notation as in Section V, namely $K_1, K_2, \cdots, K_5$. Suppose that the DM has the following set of preferences relating the importance of the criteria.

$$K_1 \sim K_2 \sim 2K_3 \sim 3K_4 \sim 4K_5,$$

i.e. the DM believes that the first criterion is just as important as the second one, is twice as important as the third etc. As discussed before we interpret the above set of preferences in terms of a linear relationship between the importance indices of a 2-additive fuzzy measure (e.g. $I_1 = I_2 = 2I_3$).

In addition, suppose the following set of interactions is given.

$$I_{12}^0 = 0.4 \quad I_{13}^0 = -0.15 \quad I_{14}^0 = 0.05$$
$$I_{23}^0 = 0.25 \quad I_{34}^0 = 0.2 \quad I_{45}^0 = -0.25$$

Remark: The formulation of the interaction index requires that $I_{ij} = I_{ji}$ for all pairs of criteria i, j. We thus use the convention that $I_{12}^0 = 0.4$ implies $I_{21}^0 = 0.4$ and vice versa. Also as was mentioned in Section IV, all interactions for which no information is given are set to 0.

Summarizing the above set of preferences as well as the well-posedness conditions for the fuzzy measure, we obtain the following optimization problem, which we solve using cvx.

$$\underset{I_i, I_{ij}}{\text{minimize}} \quad \alpha \sum_{i \in \mathcal{N}} I_i \log I_i + \sum_{i,j \in \mathcal{N}} \left(|I_{ij}| + \beta (I_{ij} - I_{ij}^0)^2 \right)$$

$$\text{subject to} \quad (I_i, I_{ij}) \in \mathcal{A}$$

The feasible set $\mathcal{A}$ is given by

$$\mathcal{A} = \left\{ \sum_{i=1}^{5} a_{ik} I_i = 0 \; \forall k, \quad \sum_{i=1}^{5} I_i = 1, \quad \text{and} \quad I_i \geq \frac{1}{2} \sum_{i \neq j} |I_{ij}| \right\}$$

and the values a_{ik} are the entries of the matrix

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -2 & 0 & 0 \\ 1 & 0 & 0 & -3 & 0 \\ 1 & 0 & 0 & 0 & -4 \end{pmatrix}$$

The obtained solution to the above problem using $\alpha = 1$ and $\beta = 5$, as discussed in the end of Section IV, is

$$\begin{aligned} &I_1 = 0.3243, \quad I_2 = 0.3243, \quad I_3 = 0.1622, \\ &I_4 = 0.1081, \quad I_5 = 0.0811 \end{aligned} \tag{6}$$

$$\begin{aligned} &I_{12} = 0.3, \quad I_{13} = -0.05, \quad I_{23} = 0.15 \\ &I_{34} = 0.0831, \quad I_{45} = -0.1331 \end{aligned} \tag{7}$$

In addition, as remarked earlier, we have that $I_{ij} = I_{ji}$ so that the above equalities are valid for the corresponding interaction indices and all other interactions are 0.

We remark that no 2-additive measure can be defined exactly on the set of preferences given by the user. The latter can be seen by first noting that the equality constraints for the importance indices together with the normalization constraint uniquely determine the the solution in (6). This implies that one of the conditions for monotonicity $I_3 \geq 1/2 \sum_i |I_{i3}^0| \iff 0.1622 \geq 0.2250$ necessarily fails.

The next step of our construction involves designing a weighing vector for an OWA operator capturing a vague statement. We consider the vague statements, which require "some" and "most" of the criteria to be satisfied. Since we are working with 5 criteria, we will interpret the two vague statements as requiring 2 and respectively 3 criteria to be satisfied. In our method for constructing the weighing vectors, described in Section V, we set $k = 2$ and $k = 3$, respectively and $\sigma = 0.10$. We obtain the following weighing vectors.

$$\mathbf{w}_1 = [0.1575 \; 0.6836 \; 0.1575 \; 0.1575 \; 0.0000] \tag{8}$$

$$\mathbf{w}_2 = [0.0013 \; 0.1573 \; 0.6827 \; 0.1573 \; 0.0013] \tag{9}$$

The choice for σ is somewhat arbitrary. In general, larger values lead to more spread out entries in the weighing vectors, whereas smaller values give more concentrated distributions. In our case, we have just over two thirds of the weight concentrated in the second and third entry of the above vectors respectively. The latter gives a sharp difference in the aggregated values (using OWA) between vectors that satisfy the vague statement poorly and those that satisfy it well.

The final step in our construction involves building the two vague fuzzy measures (corresponding to the two vague statements), using the 2-added measure obtained above as well

Criterion	$\mathbf{z_1}$	$\mathbf{z_2}$	$\mathbf{z_3}$	$\mathbf{z_4}$	$\mathbf{z_5}$	$\mathbf{z_6}$
K_1	0.6	0	0.8	0.1	0.1	0.5
K_2	0.65	0.55	0.75	0.75	0.2	0
K_3	0.65	0.85	0	0.75	0.9	1
K_4	0.8	0.9	0.1	0.75	0.9	0
K_5	0.15	1	0.8	0.5	0.9	1

TABLE III

as $\mathbf{w_1}$ and $\mathbf{w_2}$. We accomplish this as before by applying the Definition 6 in Section V.

We next consider the six vectors representing the alternatives the DM is to choose from. We denote the latter by $\mathbf{z_1}, \mathbf{z_2}, \cdots, \mathbf{z_6}$. The vectors are given in Table III.

We are next interested in the aggregated values that are obtained using the 2-additive Choquet integral, the OWA operator and the vague Choquet integral. The results are summarized in Tables IVa and IVb. The results in Table IVa are based on using $\mathbf{w_1}$ for the OWA operator and the vague Choquet integral, while those in Table IVb depend on $\mathbf{w_2}$.

TABLE IV

(a) Results using $\mathbf{w_1}$

Vector	$OWA_{\mathbf{w_1}}$	C_ν	$C_{\nu,\mathbf{w_1}}$
$\mathbf{z_1}$	0.6736	0.6402	0.6631
$\mathbf{z_2}$	0.9074	0.4154	0.6965
$\mathbf{z_3}$	0.7912	0.5771	0.7827
$\mathbf{z_4}$	0.7497	0.4543	0.721
$\mathbf{z_5}$	0.8991	0.366	0.5487
$\mathbf{z_6}$	0.9199	0.2929	0.475

(b) Results using $\mathbf{w_2}$

Vector	$OWA_{\mathbf{w_2}}$	C_ν	$C_{\nu,\mathbf{w_2}}$
$\mathbf{z_1}$	0.6417	0.6402	0.6313
$\mathbf{z_2}$	0.8097	0.4154	0.3946
$\mathbf{z_3}$	0.6547	0.5771	0.7231
$\mathbf{z_4}$	0.7098	0.4543	0.525
$\mathbf{z_5}$	0.7888	0.366	0.224
$\mathbf{z_6}$	0.5	0.2929	0.0744

Firstly, we consider the action of the Choquet integral on the given vectors. According to Table IV the highest aggregated values belong to $\mathbf{z_1}$ and $\mathbf{z_3}$, while the lowest to $\mathbf{z_5}$ and $\mathbf{z_6}$. Looking at Table III and the parameters (6) and (7) it is easy to explain the obtained results. Indeed, $\mathbf{z_1}$ satisfies well the four most important criteria, which positively interact with one another, while $\mathbf{z_3}$ satisfies very well the two most important criteria with the highest level of synergy. On the other hand, $\mathbf{z_5}$ and $\mathbf{z_6}$ mostly satisfy the least important criteria and thus receive a lower aggregated value.

The next action of interest is that of the OWA operator. From Table IVa we notice that $\mathbf{z_2}$, $\mathbf{z_5}$ and $\mathbf{z_6}$ all receive very high aggregated values, which is precisely because they have at least two entries ≥ 0.90. The vector $\mathbf{z_1}$ receives the lowest value as it contains only one relatively high entry, while $\mathbf{z_3}$

and $\mathbf{z_4}$ lie in between. The picture dramatically changes once we change the vague statement as can be seen in Table IVb. The highest aggregated values, belong to $\mathbf{z_2}$ and $\mathbf{z_5}$, while $\mathbf{z_6}$ receives the lowest rank. The vector $\mathbf{z_6}$ has exactly two very high entries, which implies that it satisfies the first vague statement well, while the second very poorly and thus the change in its rank when applying the two OWA operators is not only expected, but in fact desired. Our weighing vectors were designed to differentiate sharply between alternatives that satisfy the vague statement well and poorly, which is precisely what we observe in Table IV.

Finally, we look into the results for the vague Choquet integral. For now we restrict our attention to Table IVa. According to the vague Choquet integral, the optimal choice is $\mathbf{z_3}$, while the worst is $\mathbf{z_6}$. We remark that $\mathbf{z_3}$ satisfies well both the preferences regarding the importance and interactions as well as the vague statement as can be seen in Table IVa. It has relatively high scores with respect to $OWA_{\mathbf{w_1}}$ and C_ν and this motivates its top position when applying the vague Choquet integral. A similar argument can be made about $\mathbf{z_4}$, which again consistently received good scores with respect to the first two aggregators. Conversely, $\mathbf{z_6}$ and $\mathbf{z_1}$, which were the top-ranking vectors according to $OWA_{\mathbf{w_1}}$ and C_ν respectively, both receive a lower overall score, because they fail to satisfy some portion of the preferences well.

The situation looks similar when analyzing the results in Table IVb. We again observe that $\mathbf{z_3}$ performs well as it satisfies all types of preferences well. On the other hand we notice a shift in the ranking of $\mathbf{z_1}$, which now possesses the second overall ranking according to the vague Choquet integral. The latter is due to its better relative performance on the vague statement, compared to Table IVa. The worst performing vector is again $\mathbf{z_6}$, although we see a dramatic change in its aggregated value, compared to that in Table IVa. $\mathbf{z_6}$ poorly satisfies both types of preferences, which implies that it should receive an even lower overall score as is the case.

The three aggregation functions we considered lead to a different ordering of the alternatives, which individually and compared with respect to one another provide a reasonable ranking of the given alternatives. The vague Choquet integral clearly gives precedence to alternatives, which perform consistently well on both types of user preferences, while penalizing options that fail to satisfy either. This is the general behavior we set out to model.

B. Example II

In the previous example we looked into how the different steps of the construction of our final aggregation function acted on a set of vectors. In this example we are interested in how different vague Choquet integrals (based on different sets of preferences) rate the same set of alternatives. We can interpret this example as having several DMs with different preferences, choosing from the same options. We aim to illustrate how different preferences translate into different choices based on our model. To this end, we consider a second set of preferences for the importance and interactions of criteria.

Suppose that a different DM has the following preferences.

$$K_4 \sim 4K_1 \quad 2K_3 \sim 3K_5,$$

In addition, suppose that the he/she believes the interactions should be:

$$I_{12}^0 = 0.2 \quad I_{23}^0 = 0.35 \quad I_{25}^0 = 0.15$$
$$I_{34}^0 = 0.05; \quad I_{35}^0 = 0.1; \quad I_{45}^0 = -0.3$$

Based on the above preferences we obtain the following values for the importance and interaction indices.

$$I_1 = 0.0716 \quad I_2 = 0.217 \quad I_3 = 0.2551$$
$$I_4 = 0.2862 \quad I_5 = 0.1701 \tag{10}$$

$$I_{12} = 0.1, \quad I_{23} = 0.25, \quad I_{25} = 0.05, \quad I_{45} = -0.2 \tag{11}$$

Let us denote by ν_1 and ν_2 the 2-additive measures based on the preferences in the previous and in this example respectively. In addition, we will consider the same two vague statements as before, which are modeled via the weighing vectors $\mathbf{w_1}$ and $\mathbf{w_2}$ (given in (8) and (9)). Finally, we shall consider the ordering of the alternatives $\mathbf{z_1}, \mathbf{z_2}, \cdots, \mathbf{z_6}$, induced by the four vague Choquet integrals - $C_{\nu_1,\mathbf{w_1}}, C_{\nu_1,\mathbf{w_2}}, C_{\nu_2,\mathbf{w_1}}, C_{\nu_2,\mathbf{w_2}}$. The aggregated values are provided in Table Va and for clarity we present the ordering of the vectors in Table Vb.

TABLE V

(a) Aggregated values

Vector	$C_{\nu_1,\mathbf{w_1}}$	$C_{\nu_1,\mathbf{w_2}}$	$C_{\nu_2,\mathbf{w_1}}$	$C_{\nu_2,\mathbf{w_2}}$
$\mathbf{z_1}$	0.6631	0.6313	0.7691	0.6712
$\mathbf{z_2}$	0.6965	0.3946	0.915	0.7795
$\mathbf{z_3}$	0.7827	0.7231	0.6831	0.2661
$\mathbf{z_4}$	0.721	0.525	0.75	0.7461
$\mathbf{z_5}$	0.5487	0.224	0.8778	0.6969
$\mathbf{z_6}$	0.475	0.0744	0.7934	0.1477

(b) Rank of vectors

Vector	$C_{\nu_1,\mathbf{w_1}}$	$C_{\nu_1,\mathbf{w_2}}$	$C_{\nu_2,\mathbf{w_1}}$	$C_{\nu_2,\mathbf{w_2}}$
$\mathbf{z_1}$	3	3	2	2
$\mathbf{z_2}$	4	1	5	4
$\mathbf{z_3}$	2	4	6	5
$\mathbf{z_4}$	1	2	1	1
$\mathbf{z_5}$	5	5	4	3
$\mathbf{z_6}$	6	6	3	6

As can be seen from Table Vb the different sets of preferences naturally lead to different ranking of the alternatives. When the preferences of the DM are modeled using ν_1, we see that for both vague statements $\mathbf{z_3}$ is the optimal choice. In particular, $\mathbf{z_3}$ has three high entries and thus satisfies both vague statements well. A similar situation can be seen for $\mathbf{z_2}$, when the 2-additive measure used is ν_2.

In order to see how vague statements change the rank of an option we only need to consider $\mathbf{z_6}$. $\mathbf{z_6}$ has exactly two high entries and thus it is a good representative of a vector that satisfies only one of the vague statements well. And as

can be seen in Table Vb, when the 2-additive measure is ν_2, the ranking of $\mathbf{z_6}$ drastically changes depending on the given vague statement. Such a situation is not observed when ν_1 is used, because as seen in Table IVa, $\mathbf{z_6}$ fails to satisfy the preferences modeled by ν_1, and its final aggregated score is lowest when either vague statement is used.

Finally, we remark on the effect of the 2-additive measure, reflecting the importance and interactions. When ν_1 is used $\mathbf{z_3}$ is the best option, as it has high entries in the most important and most synergetic two criteria. That is no longer the case when the preferences are modeled by ν_2, and as can be seen in Table Vb the rank of $\mathbf{z_3}$ drops to the very bottom.

VII. CONCLUSION

We presented a novel and efficient method for constructing an aggregation function capturing two types of information: vague statements as used in the OWA and importance/interactions of criteria as used in the Choquet integral. The proposed aggregation function, called a vague Choquet integral, successfully combines two types of information that were previously unintegrated. The proposed model is built automatically and without the need of any data learning. The vague fuzzy measure, underlying our final model, is constructed by combining a 2-additive fuzzy measure and a weighing vector capturing partial user information. The weighing vector reflects the given vague statement and is constructed using weight generating functions based on the truncated Gaussian distribution. The 2-additive fuzzy measure reflects importance and interactions of criteria and is obtained by minimizing an L^1 energy function. The solution to the optimization problem is a 2-additive fuzzy measure, whose interaction profile is sparse. To our knowledge this paper presents the first application of L^1 optimization in multi-criteria decision-making problems, which provides an automatic construction of a fuzzy measure.

The proposed model was shown to better incorporate a complex set of user preferences than the OWA or the Choquet integral separately. Some future research includes finding alternative and better ways to interpret user preferences as constraints within the optimization problem. In addition, it would be interesting to consider a MACBETH-type approach for reflecting the contribution of each preference to the final aggregated value. This would lead to a better understanding of the presented models and improve their applicability to real world problems.

REFERENCES

[1] G. Beliakov, A. Pradera and T. Calvo. *Aggregation Functions: A Guide for Practitioners*, Springer, Berlin, New York (2007).
[2] S. Boyd and M. Grant. *CVX: Matlab Software for Disciplined Convex Programming v. 1.21*, http://cvxr.com/cvx, (2011)

[3] E.J. Candes, J. Romberg and T. Tao. *Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information*, IEEE Transactions on Information Theory, Vol. 52, Issue 2 (2006), pp.489-509

[4] E.J. Candes, *The Restricted Isometry Property and its Implications for Compressed Sensing*, Comptes Rendus Mathematique, Vol. 346, Issues 9-10 (2008), pp. 589-592

[5] E. Candes et. al. *Robust Principal Component Analysis?*, arXiv:0912.3559v1[cs.IT] (2009)

[6] A. Chambolle. *An Algorithm for Total Variation Minimization and Applications*, J. of Mathematical Imaging and Vision, Vol. 20, Numbers 1-2 (2004), pp. 89-97

[7] S.-B. Cho. *Fuzzy Aggregation of Modular Neural Networks with Ordered Weighted Averaging Operators*, Int. J. of Approximate Reasoning, Vol. 13, Issue 4 (1995), pp.359-375

[8] G. Choquet. *Theory of Capacities*, Annales de I'institut Fourer, 5 (1954), pp. 131-295

[9] D. Diakoulaki, C. Antunes and A. Martins. *MCDA in Energy Planning*, Int. Series in Operations Research and Management Science, Vol. 78, Issue 7 (2005), pp. 859-890

[10] D.L. Donoho. *Compressed Sensing*, IEEE Transactions on Information Theory, Vol. 52, Issue 4 (2006), pp. 1289-1306

[11] J. Figueira, S. Greco and B. Roy. *ELECTRE Methods with Interaction between Criteria: An Extension of the Concordance Index*, European J. of Operational Research, Vol. 199, Issue 2 (2009), pp. 478-495

[12] D. Filev and R. Yager. *Analytic Properties of Maximal Entropy OWA Operators*, Information Sciences, Vol. 85, Issues 1-3 (1995), pp. 11-27

[13] T. Goldstein and S. Osher. *The Split Bregman Algorithm for L1 Regularized Problems*, UCLA CAM Report 08-29, UCLA, Los Angeles (2008)

[14] M. Grabisch and J.-M. Nicolas. *Classification by Fuzzy Integral: Performance and Tests*, Fuzzy Sets and Systems, Vol. 65, Issues 2-3 (1994), pp.255-271

[15] M. Grabisch. *Fuzzy Integral in Multicriteria Decision Making*, Fuzzy Sets and Systems 69 (1995), pp. 279-298

[16] M. Grabisch. *The Application of Fuzzy Integrals in Multicriteria Decision Making*, European J. of Operational Research 89 (1996), pp. 445-456

[17] M. Grabisch. *k-order Additive Fuzzy Measures*, 6th Int. Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Granada, Spain (1996), pp. 1345-1350

[18] M. Grabisch. *k-order Additive Discrete Fuzzy Measures and their Representation*, Fuzzy Sets and Systems, Vol. 92, Issue 2 (1997), pp. 167-189

[19] M. Grabisch. *Alternative Representations of Discrete Fuzzy Measures for Decision Making*, Int. J. of Uncertainty, Fuzzyness, Knowledge-Based Systems, Vol. 5 (1997), pp. 587-607

[20] M. Grabisch et. al. *Subjective Evaluation of Discomfort in Sitting Position*, Fuzzy Optimization and Decision Making, Vol. 1, Number 3 (2002), pp. 287-312

[21] M. Grabisch and Ch. Labreuche. *A Decade of Application of the Choquet and Sugeno Integrals in Multi-Criteria Decision Aid*, Annals of Operations Research, Vol. 175 (2010), pp. 247-286

[22] F. Herrera, E. Herrera-Viedma and J.L. Verdegay. *A sequential Selection Process in Group Decision Making with a Linguistic Assessment Approach*, Information Sciences, Vol. 85, Issue 4 (1995), pp.223-239

[23] F. Herrera, E. Herrera-Viedma and J.L. Verdegay. *A Model of Consensus in Group decision making Under Linguistic Assessments*, Fuzzy Sets and Systems, Vol. 78, Issue 1 (1996), pp.73-87

[24] H. Jiang and R. Eastman. *Application of Fuzzy Measures in Multicriteria Evaluation in GIS*, Int. J. of Geographical Information Science, Vol. 14, Issue 2 (2000), pp.173-184

[25] I. Kojadinovic. *Minimum Variance Capacity Identification*, European J. of Operational Research, Vol. 177, Issue 1 (2007), pp.498-514

[26] D. Liginlal, T. Ow. *Modeling Attitude to Risk in Human Decision Processes: An Application of Fuzzy Measures*, Fuzzy Sets and Systems, Vol. 157, Issue 23 (2006), pp. 3040-3054

[27] M. Lustig et. al. *Faster Imaging with Randomly Pertrubed, Undersampled Spirals and —L—1 Reconstruction*, In Proc. Of the ISMRM '05 (2005).

[28] M. Lustig, D. Donoho and J.M. Pauly. *Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging*, Magnetic Resonance in Medicine, Vol. 58, Issue 6 (2007), pp. 1182-1195

[29] J.-L. Marichal and M. Roubens. *Entropy of Discrete Fuzzy Measures*, Int. J. of Uncertainty, Fuzziness, and Knowledge Based Systems 8 (6) (2000), pp. 625-640

[30] J.-L. Marichal and M. Roubens. *Determination of Weights of Interacting Criteria from a Reference Set*, European J. of Operational Research, Vol. 124, Issue 3 (2000), pp. 641-650

[31] J.-L. Marichal, P. Meyer and M. Roubens, *Sorting Multi-Attribute Alternatives: The TOMASO Method*, Computers and Operations Research, Vol. 32, Issue 4 (2005), pp. 861-877

[32] P. Meyer and M. Roubens. *On the Use of the Choquet Integral with Fuzzy Numbers in Multiple Criteria Decision Aiding*, Fuzzy Sets and Systems, Vol. 157, Issue 7 (2006), pp. 927-938

[33] P. Miranda and M. Grabisch. *Optimization Issues for Fuzzy Measures*, Int. J. of Uncertainty, Fuzziness, and Knowledge Based Systems 7 (1999), pp. 545-560

[34] E.-M. Nordström, L. Eriksson and K. Öhman. *Integrating Multiple Criteria Decision Analysis in Participatory Forest Planning: Experience form a Case Study in Northern Sweden*, Forest Policy and Economics, Vol. 12, Issue 8 (2010), pp. 562-574

[35] C. Rinner and J. Malczewski. *Web-enabled Spatial Decision Analysis Using Ordered Weighted Averaging(OWA)*, J. of Geographical Systems, Vol. 4, Issue 4 (2002), pp. 385-403

[36] L. Rudin, S. Osher and E. Fatemi. *Nonlinear Total Variation Based Noise Removal Algorithms*, Physica D: Nonlinear Phenomena, Vol. 60, Issues 1-4 (1992), pp. 259-268

[37] H. Schaeffer and S. Osher. *A Low Patch-Rank Interpretation of Texture*, CAM Report (Nov. 2011)

[38] M. Sugeno. *Theory of Fuzzy Integrals and its Applications*, PhD Thesis, Tokyo Institute of Technology (1974)

[39] C. Tan and X. Chen. *Intuitionistic Fuzzy Choquet Integral Operator for Multi-Criteria Decision Making*, Expert Systems with Applications, Vol. 37, Issue 1 (2010), pp.149-157

[40] O. Thonnard. *A Multi-Criteria Clustering Approach to Support Attack Attribution in Cyberspace*, PhD Thesis, Ecole Doctorale d'Informatique, Telecommunications et Electronique de Paris (2010)

[41] O. Thonnard, W. Mees and M. Dacier. *On a Multicriteria Clustering Approach for Attack Attribution*, SIGKDD Explorations (2010), pp.11-20

[42] V. Torra. *Weighted OWA Operators for Synthesis of Information*, Proceedings of the Fifth IEEE Int. Conference on Fuzzy Systems, Vol. 2 (1996), pp. 966-971

[43] V. Torra. *The Weighted OWA Operator*, Int. J. of Intelligent Systems, Vol. 12, Issue 2 (1997), pp. 153-166

[44] A. Volkova, E. Latõsov and A. Siirde. *Use of Multi-Criteria Decision Analysis for Choosing an Optimal Location for a Wood Fuel Based Cogeneration Plant: A Case Study in Estonia*, Selected Topics in Energy, Environment, Sustainable Development and Landscaping: WSEAS Press (2010), pp. 88-94

[45] J.-J. Wang et. al. *Review on Multi-Criteria Decision Analysis Aid in Sustainable Energy Decision-Making*, Renewable and Sustainable Energy Reviews, Vol. 13, Issue 9 (2009), pp. 2263-2278

[46] R. Yager. *On Ordered Weighted Averaging Aggregation Operators in Multicriteria Decision Making*, IEEE Trans. Systems Man Cybernet, Vol. 18, Issue 1 (1988), pp. 183-190

[47] R. Yager. *On a Semantics for Neural Networks Based on Fuzzy Quantifiers*, Int. J. of Intelligent Systems, Vol. 7, Issue 8 (1992), pp. 765-786

[48] R. Yager. *Quantifier Guided Aggregation Using OWA Operators*, Int. J. of Intelligent Systems, Vol. 11, Issue 1 (1996), pp. 49-73

[49] C. Zopounidis. *Multicriteria Decision Aid in Financial Management*, European J. of Operational Research, Vol. 119, Issue 2 (1999), pp. 404-415

[50] C. Zopounidis and M. Doumpos. *Multi-Criteria Decision Aid in Financial Decision Making: Methodologies and Literature Review*, J. of Multi-Criteria Decision Analysis, Vol. 11, Issues 4-5 (2002), pp. 167-186

[51] MATLAB version R2012a, Natick, Massachusetts: The MathWorks Inc., (2012)