



**HAL**  
open science

## Multicriteria Decision Making

Christophe Gonzales, Patrice Perny

► **To cite this version:**

Christophe Gonzales, Patrice Perny. Multicriteria Decision Making. Marquis, Pierre; Papini, Odile; Prade, Henri. A Guided Tour of Artificial Intelligence Research, I, Springer, pp.519-548, 2020, Knowledge Representation, Reasoning and Learning, 978-3-030-06164-7. 10.1007/978-3-030-06164-7\_16 . hal-02860326

**HAL Id: hal-02860326**

**<https://hal.science/hal-02860326>**

Submitted on 12 Mar 2021

**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.

# Multicriteria Decision Making

Christophe Gonzales<sup>1</sup> and Patrice Perny<sup>2</sup>

**Abstract** This chapter aims to present the main models used for preference aggregation and decision support in a unified framework. After recalling the definition of a multicriteria decision making problem, we distinguish two approaches for preference aggregation: *the compare then aggregate* approach (denoted CA) and the *aggregate then compare* approach (denoted AC). We first present some procedures allowing the construction of an overall preference relation (e.g., a dominance or concordance relation) from several binary relations. Then we consider the AC approach and present some scalarizing functions allowing the definition of an overall score from partial numerical evaluations. In particular we review the min, Tchebycheff, OWA, WOWA aggregators and Choquet and Sugeno integrals.

## 1 Introduction

Taking into account multiple and conflicting points of view in the analysis of preferences and studying the properties of preference aggregation procedures is quite old. Long before the birth of multicriteria optimization, collective decision-making problems and aggregation of preferences were addressed through the theory of voting, as can be seen in the writings of Borda [Borda, 1781] and Condorcet [Marquis de Condorcet, 1785]; these topics remained active until today, giving birth to the theory of social choice [Arrow, 1951; Sen, 1986a]. In economics, the account of multiple criteria to explain rational behaviors dates back to the 1900s, notably with the works of Pareto [Pareto, 1906]. The consideration of multiple objectives in mathematical programming was introduced in the middle of the 20th century with the goal-programming [Charnes et al., 1955]. This work was then developed in the 1970s

---

<sup>1</sup>Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France,  
e-mail: christophe.gonzales@lis-lab.fr

<sup>2</sup>Sorbonne Université, CNRS, LIP6, Paris, France,  
e-mail: patrice.perny@lip6.fr

under the impetus of Geoffrion (see for example [Geoffrion et al., 1973]). The first international conference dedicated to multicriteria decision-making took place in 1973 in South Carolina. This activity continued in Europe with the first Euro Working Group on multi-criteria decision-making led by Bernard Roy who appeared in 1975. The work on multiobjective combinatorial optimization followed a few years later and the first international conference on the subject (MOPGP) was established in 1994. In Artificial intelligence, taking into account multiple objectives in problem solving also appears in the same period, in particular in multiobjective state-space search [Stewart and White III, 1991; Laumanns et al., 2002], multicriteria planning problems [White, 1982; Gábor et al., 1998], and problems of satisfaction of flexible constraints [Schiex et al., 1995]. Concerning modeling and aggregation of preferences, the work in AI is distinguished by the emphasis on qualitative models and, more recently, on the aspects of representation and automatic learning of preferences. The latter two aspects go beyond the scope of this chapter, which focuses on the main models used for preference aggregation.

## 2 Multicriteria Decision Problems

A *multicriteria* decision problem is characterized by the explicit consideration of several objectives to be optimized simultaneously in the analysis of preferences, the comparison of solutions and the determination of the optimal solution(s). To formally introduce a multicriteria decision problem, we first define a set  $A$  of alternatives (potential actions, feasible solutions, candidates) that can be given explicitly (for example by listing the solutions considered) or implicitly (for example by specifying a set of constraints or properties that the solutions must satisfy). In all cases, the  $A$  set defines the solutions on which one wishes to make the decision analysis, that is to say that they are the only ones available, the only ones realizable or the only ones admissible. We then introduce a finite set of criteria  $N = \{1, \dots, n\}$  taking the form of functions  $f_i, i \in N$ , modeling the objectives of the decision maker. For every  $x \in A$  and every  $i \in N$ , we call *performance* of  $x$  on criterion  $i$  the quantity  $f_i(x)$  reflecting the value of  $a$  with respect to criterion  $i$ . This formalism also applies to problems of collective decision-making. In such problems,  $N$  represents a set of agents and the performance  $f_i(x)$  then represents the utility of the solution  $x$  from the point of view of agent  $i$ . The criteria  $f_i$  are defined on  $A$  and respectively valued in an ordered set  $X_i, i \in N$ .

The set  $\mathcal{X} = X_1 \times \dots \times X_n$  which constitutes a new description space of the alternatives, is called *the space of criteria*. To simplify the notations, we will set  $x_i = f_i(x)$  for all  $x \in A$  and all  $i \in N$ . Any solution  $x$  is therefore represented in  $\mathcal{X}$  by a vector  $(x_1, \dots, x_n)$ . We will assume that  $X_i \subseteq \mathbb{R}$  for all  $i \in N$  and therefore that  $\mathcal{X} \subseteq \mathbb{R}^n$ . To simplify the presentation, it will also be assumed that functions  $f_i$  are to be maximized (this is not restrictive because we can change the sign of evaluations to pass from minimization to maximization). Sometimes performances have no cardinal meaning and only their order matters to state preferences. In other

cases they may represent a cardinal utility (uniquely defined up to a positive affine transformation) or even an absolute evaluation reflecting the intrinsic attractiveness of a solution on each of the criteria considered.

Multicriteria decision problems vary depending on the question asked. We can distinguish *choice* problems where we try to determine the best solutions, *ranking* problems where the aim is to order, at least partially, the solutions according to their relative merit, or *sorting* problems where one seeks to assign the solutions to pre-defined categories according to their intrinsic value. To summarize, a multicriteria decision problem can always be formally characterized by a triple of the form:

$$(A, \{f_1, \dots, f_n\}, Q)$$

where  $Q \in \{\text{choice, ranking, sorting}\}$  is the question formulated. The choice problem is the one most often encountered in the theory and practice of multicriteria analysis. It aims to find a solution that optimizes the different criteria as well as possible, or to find a subset of solutions, as small as possible, containing the best solutions. Formulated in this way, multicriteria optimization is an ill-posed problem. Indeed, due to the potentially conflicting nature of the criteria, there is generally no solution optimizing all the criteria simultaneously. The only natural preference weak-order that can be built on  $A$  without adding preferential information to the description of the problem is in fact the so-called *weak Pareto dominance* relation denoted  $\succsim_P$  and is defined as follows:

$$x \succsim_P y \text{ iff } [\forall i \in N, x_i \geq y_i]$$

Relation  $\succsim_P$  is a weak partial order on  $A$  (i.e. a reflexive and antisymmetric binary relation, transitive but not complete) and it generally leaves many pairs of solutions incomparable. It is enough that a solution  $x$  is better than a solution  $y$  on one criterion and that it is the opposite on another criterion so that we can no longer compare them. To circumvent this difficulty, one generally seeks to construct a richer and more discriminating preference relation on the set of alternatives. This relation will be denoted here  $\succsim$  with the convention that  $x \succsim y$  means that  $x$  is judged at least as good as  $y$  given the performance vectors  $(x_1, \dots, x_n)$  and  $(y_1, \dots, y_n)$ . Obviously one will generally be interested in constructing a transitive preference relation  $\succsim$  such that  $x \succsim_P y \Rightarrow x \succsim y$  for all  $x, y \in A$ , thus refining the weak Pareto dominance.

We can also define a strict preference relation  $\succ$  as the asymmetric part of  $\succsim$ . We obtain  $x \succ y$  iff  $[x \succsim y \text{ and not}(y \succsim x)]$ . For example, the asymmetric part of relation  $\succsim_P$  is the so-called *Pareto dominance* denoted  $\succ_P$ . We have:

$$x \succ_P y \text{ iff } \begin{cases} \forall i \in N, x_i \geq y_i \\ \exists k \in N, x_k > y_k \end{cases}$$

Given a strict preference relation  $\succ$  defined on  $A$ , the non-dominated solutions of  $A$  are formally defined as follows:  $ND(A, \succ) = \{x \in A : \forall y \in A, \text{non}(y \succ x)\}$ . This set is non-empty as soon as  $\succ$  is transitive. For example, the set  $ND(A, \succ_P)$  is never empty; this is the set of Pareto-optimal solutions also known as the Pareto set. If,

as suggested above, we work with a preference relation  $\succ$  which refines the Pareto dominance  $\succ_P$ , the set  $ND(A, \succ)$  will therefore be a subset of the Pareto set.

The overall weak preference relation  $\succsim$  must be constructed from the  $n$  criteria representing the different points of view considered relevant in the analysis, taking into account how the decision-maker wants to define the resultant of potentially conflicting criteria. This is the multicriteria aggregation phase that we present in the next subsection.

## 2.1 Preference Aggregation

The *preference aggregation* problem consists of synthesizing information that reflects different aspects or points of view (e.g., performance indices, utilities, preferences), sometimes conflicting, on a same set of alternatives. It is critically important in many procedures used for assessment, comparison or classification in multicriteria decision support. Whether it is a problem of choice, ranking or sorting, the central question is always a problem of comparison. Thus, in a problem of choice, the identification of the best candidate needs to be able to compare it to all others; in a ranking problem, we need to compare any pair of alternatives; in sorting problems, assigning a solution to a category is often done by comparing the solution to a reference vector. In a multicriteria decision problem, the comparison of two solutions is performed on the basis of their respective performance vectors. For this purpose, one necessarily resorts to an *aggregation rule* to construct the overall preference relation  $\succsim$ . Aggregation rules can formally be introduced as follows:

**Definition 1.** An *aggregation rule* is a function that defines the preference  $x \succsim y$  for any pair of alternatives  $(x, y)$  in  $A \times A$  from performance vectors  $(x_1, \dots, x_n)$  and  $(y_1, \dots, y_n)$  as follows:

$$x \succsim y \text{ iff } h(x_1, \dots, x_n, y_1, \dots, y_n) \geq 0 \quad (1)$$

where  $h$  is a real-valued function defined on  $\mathbb{R}^{2n}$ , non-decreasing in the  $n$  first arguments and non-increasing in the  $n$  last arguments such that  $h(x, x) \geq 0$  for all  $x \in \mathbb{R}^n$  (which enforces the reflexivity of  $\succsim$  and the compatibility with the weak Pareto-dominance).

The  $h$  function which tests the preference of  $x$  over  $y$  performs on the one hand the *aggregation* of performances  $x_i$  and  $y_i$ ,  $i = 1, \dots, N$  and, on the other hand, the *comparison* of solutions  $x$  and  $y$  through their performance vectors. Generally, these two steps (aggregation and comparison) are clearly distinguished and  $h$  is then defined as the combination of an aggregation function (also known as a scalarizing function)  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  allowing the synthesis of a vector of  $n$  performances in one scalar, and a performance comparison function  $\phi : \mathbb{R}^2 \rightarrow \mathbb{R}$  which compares two performances. We can thus distinguish two different modes of operation giving rise to two distinct approaches to multicriteria aggregation.

### The “Aggregate then Compare” approach (AC)

It consists of summarizing the value of any solution  $x$  by an overall score  $\psi(x)$  calculated from its performance vector. This score is intended to summarize the overall value of the action and serves as a basis for the multicriteria comparison of solutions. This way of evaluating and comparing vectors of grades is very widespread, for example in the academic world where the comparison of two students is based on the average of their marks. The general form of the decision rules under the AC approach is as follows:

$$x \succsim y \text{ iff } \phi(\psi(x_1, \dots, x_n), \psi(y_1, \dots, y_n)) \geq 0 \quad (2)$$

where  $\psi$  is a non-decreasing function of its arguments. On the other hand, function  $\phi$  allows the comparison of  $x$  and  $y$  on the basis of  $\psi(x)$  and  $\psi(y)$ . The most common choice for  $\phi$  is  $\phi(x, y) = x - y$ . In this case,  $x \succsim y$  holds when  $\psi(x) \geq \psi(y)$ .

*Example 1.* The Nash product often used in Game Theory is an aggregation function leading to the following preference relation:

$$x \succsim y \text{ iff } \prod_{i=1}^n x_i \geq \prod_{i=1}^n y_i$$

This is clearly an instance of the AC approach where  $\psi$  is the product function and  $\phi(x, y) = x - y$ .

### The “Compare then Aggregate” Approach (CA)

It consists of comparing, criterion by criterion, the performances of the alternatives and then to aggregate these comparisons. Thus, for each pair  $(x, y)$  and each criterion  $i$ , one can define a binary index of partial preference  $\phi_i(x, y)$  where  $\phi_i$  is an increasing function of  $x_i$ , decreasing of  $y_i$ . The preference  $x \succsim y$  is then defined by aggregating partial preference indices. Formally, we have:

$$x \succsim y \text{ iff } \psi(\phi_1(x, y), \dots, \phi_n(x, y)) \geq 0 \quad (3)$$

Generally in this approach, each  $\phi_i$  function is used to compare the performances of two alternatives on the same criterion (criterion  $i$ ). There are however a few exceptions in the case of criteria sharing the same valuation scale (e.g., the Lorenz dominance relation introduced later in the chapter). In this latter case,  $\phi_i$  is used to compare two performances associated with different criteria. In the CA approach, one can use the same aggregation functions  $\psi$  as for the AC approach but it is used for the aggregation of partial preference indices  $\phi_i(x, y)$  and not for aggregating the performances themselves. It must therefore be assumed that one can compare quantities of type  $\phi_i(x, y)$  and  $\phi_k(x, y)$  for  $i \neq k$  but it is not necessary to assume that we can compare the performances of an alternative on different criteria. If  $\phi_i(x, y)$

depends only on  $x_i$  and  $y_i$ , then it defines a preference relation on  $A$  restricted to criterion  $i$ . For example, we can consider preorders constructed from performances by setting  $x \succsim_i y$  if and only if  $\phi_i(x, y) \geq 0$  with:

$$\phi_i(x, y) = \begin{cases} 1 & \text{if } x_i > y_i \\ 0 & \text{if } x_i = y_i \\ -1 & \text{if } x_i < y_i \end{cases} \quad (4)$$

It is also possible to use, for  $\phi_i$ , functions with thresholds such as  $\phi_i(x, y) = x - y + q_i$  where  $q_i$  is a positive quantity representing an indifference threshold, i.e., the biggest difference that is not significant of a preference. In this case, if we set  $x \succsim_i y$  iff  $\phi_i(x, y) \geq 0$ , we obtain a semi-order structure<sup>1</sup> well known in preference modeling. Other choices are possible for  $\phi_i$ , leading to more general ordinal structures defined from numerical evaluations, see [Roubens and Vincke, 1985; Pirlot and Vincke, 1997]. Alternatively, the  $\phi_i(x, y)$  indices can be used to represent preference intensities, monotonically increasing with differences of type  $x_i - y_i$ ; this amounts to defining a fuzzy preference relation  $\succsim_i$  for every criterion  $i \in N$  [Perny and Roy, 1992; Fodor and Roubens, 1994]). In all these cases, we see that the CA approach amounts to construct  $n$  preference relations (one per criterion) and then to aggregate these relations. By way of illustration, let us give the following example:

*Example 2.* The lexicographic aggregation is characterized by the following definition of the overall preference:

$$x \succ y \text{ iff } \exists k \in N, \begin{cases} x_k > y_k \\ \forall i < k, x_i = y_i \end{cases}$$

this is clearly an instance of the CA approach with  $\psi(z_1, \dots, z_n) = \sum_{i=1}^n 2^{n-i} z_i$  and  $\phi_i$  defined as in Equation (4).

### Comparison of AC and CA

Both approaches AC and CA are represented in the following diagram showing the two possible paths to decide whether  $x$  is preferred to  $y$  from two performance vectors  $x = (x_1, \dots, x_n)$  and  $y = (y_1, \dots, y_n)$ :

$$\begin{array}{ccc} (x_1, \dots, x_n), (y_1, \dots, y_n) & \xrightarrow{\psi} & \psi(x), \psi(y) \\ \phi_i \downarrow & & \downarrow \phi \\ \phi_1(x, y), \dots, \phi_n(x, y) & \xrightarrow{\psi} & x \succsim y? \end{array}$$

<sup>1</sup> We recall that a semi-order is a complete, Ferrers and semi-transitive binary relation (see [Pirlot and Vincke, 1997]). In a semi-order  $\succsim_i$  defined with threshold  $q_i$ , we have  $x \succ_i y$  if  $x_i - y_i > q_i$  and  $x \sim_i y$  if  $|x_i - y_i| \leq q_i$ .

The fundamental difference between AC and CA lies in the fact that the aggregation does not involve the same objects due to the inversion of the order of execution of the aggregation and comparison operations. In AC, function  $\psi$  aggregates the performances  $x_i$  on the one hand and the performances  $y_i$  on the other hand whereas in CA we aggregate the preference indices  $\phi_i(x, y)$ . It should be emphasized that this distinction concerns the process of computing  $h$  in Equation (1) and thus remains essentially formal. There are indeed some aggregation rules whose function  $h$  can be written either in one or the other of the two forms distinguished. For example, this is the case of preferences defined from linear aggregations of performances as follows:

$$x \succsim y \text{ iff } \sum_{i=1}^n x_i \geq \sum_{i=1}^n y_i \quad (5)$$

In this case, one naturally recognizes an instance of the AC approach:

$$\begin{aligned} h(x_1, \dots, x_n, y_1, \dots, y_n) &= \phi(\Psi(x_1, \dots, x_n), \Psi(y_1, \dots, y_n)) \\ \Psi(z_1, \dots, z_n) &= \sum_{j=1}^n z_j \\ \phi(u, v) &= u - v \end{aligned}$$

Nevertheless, using the same  $\psi$  function, one may also consider that it is an instance of the CA approach by setting:

$$\begin{aligned} h(x_1, \dots, x_n, y_1, \dots, y_n) &= \Psi(\phi_1(x, y), \dots, \phi_n(x, y)) \\ \phi_i(x, y) &= x_i - y_i \end{aligned}$$

Despite this non-empty intersection due to some singular cases, the distinction between the AC and CA approaches is important for structuring the field of multicriteria aggregation rules (but also for transposing them into the field of decision under uncertainty) and to differentiate between the main benefits and disadvantages of each of these approaches.

The choice of the AC approach based on the construction of a “scalarizing” function  $\psi$  often seduces by its operational simplicity and its intuitive appearance. When the preferential information is sufficiently rich to enable the construction of the overall evaluation function  $\psi$ , the multicriteria decision problem reduces to the problem of optimizing function  $\psi$ . However, it should be stressed that the AC approach requires a particularly rich information. We need to know how to commensurate partial evaluations on different criteria, what is the importance of every criteria and group of criteria, and how they interact in the definition of preferences. Moreover, the preference  $x \succsim y$  is often defined using  $\phi(\psi(x), \psi(y)) = \psi(x) - \psi(y)$  which simply amounts to comparing the two values  $\psi(x)$  and  $\psi(y)$ . This mode of comparison therefore presupposes *a priori* that all solutions are reducible to scalar values and are comparable, and that preferences are transitive. From a descriptive point of view, this hypothesis is often debatable, given the necessarily imperfect information available, the heterogeneity of the performances and also the existence possible con-

flicts between the criteria considered, making it difficult to compare certain pairs of solutions.

The CA approach, on the other hand, is particularly suited to multicriteria aggregation when there is insufficient information to compare the performances issued from distinct criteria and/or when certain criteria are quantitative and others are qualitative, or when we want to relax the assumption that all the alternatives are comparable. A binary relation  $\succsim_i$  is constructed using  $\phi_i$  for every  $i \in N$  and then we have to aggregate the  $(\succsim_1, \dots, \succsim_n)$ . This ordinal aggregation problem is generally quite difficult to solve and a number of negative results have shown the prescriptive limits of decision rules based on an ordinal aggregation (for a compilation of the main results see [Sen, 1986b]). In particular, a method of aggregation defined by Equations (3) and (4) does not guarantee any transitivity property for  $\succsim$  except in some very restrictive particular cases such as the lexicographic preferences introduced in Example 2. If we consider, for example, the majority rule that appears naturally as an instance of the CA approach obtained by setting  $\psi(z_1, \dots, z_n) = \sum_{i=1}^n z_i$  and  $\phi_i(x, y)$  defined as in Equation (4), a non-transitive preference is easily obtained as shown in the following example:

*Example 3.* Let us consider a problem with three criteria to be maximized and three alternatives  $x = (3, 1, 2)$ ,  $y = (2, 3, 1)$  and  $z = (1, 2, 3)$ . We have:  $\phi_1(x, y) = \phi_3(x, y) = 1$ ,  $\phi_2(x, y) = -1$  and therefore  $x \succ y$ . On the other hand we have:  $\phi_1(y, z) = \phi_2(y, z) = 1$ ,  $\phi_3(y, z) = -1$  and therefore  $y \succ z$ . Finally we have:  $\phi_2(z, x) = \phi_3(z, x) = 1$ ,  $\phi_1(z, x) = -1$  and therefore  $z \succ x$ . This intransitivity of the strict majority rule is well known in voting theory under the name of Condorcet paradox.

Despite the descriptive appeal of the majority rule, this lack of transitivity is rather problematic for determining a ranking of solutions or even a choice and an additional exploitation phase is then necessary. For this reason, the CA approach is mainly used for decision problems involving only a finite and small set of alternatives. The next section is intended to introduce some standard decision models falling in the CA approach and, when necessary, some techniques for making a choice or ranking from a non-transitive relation constructed with this approach. We will review the main models of the AC approach in the following section.

### 3 Decision Models in the CA approach

Below, we introduce different binary relations  $\succsim$  to compare the solutions of  $A$  on the basis of their performance vectors. Let us first introduce *dominance* relations and then *outranking* relations resulting from a concordance rule.

### 3.1 Dominance Relations

We have previously introduced the notions of weak Pareto dominance which is a basic preference relation. It is not very discriminating but can be enriched in various ways. We now present some richer dominance relations often used in decision theory. Most of these dominance relations are obvious instances of the CA approach. We will therefore not specify the functions  $\psi$  and  $\phi_i$  characteristic of the CA approach, except in the few cases where membership to CA is less straightforward.

#### Oligarchic Dominance

An *Oligarchic weak-dominance* is a transitive preference relation that concentrates the decisive power on a subset of criteria  $O \subseteq N$ , namely the *Oligarchy*. It is formally defined as follows:

$$x \succsim y \text{ iff } \forall i \in O, x_i \geq y_i \quad (6)$$

When  $O = N$  we obtain the weak Pareto dominance introduced before. When  $O$  only contains some of the criteria, the dominance defined by Equation (6) is all the more discriminating than the cardinal of  $O$  is reduced. When  $O$  is reduced to a singleton we obtain a *dictatorial* aggregation rule. A refinement of this dictatorial procedure is given by the lexicographic procedure introduced in Example 2.

#### $\varepsilon$ -dominance

An interesting weakening of the weak Pareto dominance is the  $\varepsilon$ -dominance defined as follows:

$$x \succsim_{\varepsilon} y \text{ iff } \forall i \in N, (1 + \varepsilon)x_i \geq y_i$$

for some arbitrarily small  $\varepsilon > 0$ . This relation is not transitive but it enables to “cover” the entire set of feasible alternatives with fewer solutions than needed for the weak Pareto dominance, as shown in the following example:

*Example 4.* Consider a tri-criteria problem with 4 feasible solutions  $x = (10, 10, 10)$ ,  $y = (11, 5, 10)$ ,  $z = (10, 2, 11)$ ,  $w = (4, 10, 3)$ . In this example, the solutions  $x, y, z$  are Pareto-optimal while  $w$  is Pareto-dominated by  $x$ . If we consider  $\varepsilon = 0.1$  then we can check that  $x \succsim_{\varepsilon} y$ ,  $x \succsim_{\varepsilon} z$  and  $x \succsim_{\varepsilon} w$ . We thus observe that  $x$  is at least as good as all the other solutions and covers by itself all the solutions under consideration. This can be an argument for choosing this solution. One can even check here that  $x$  strictly  $\varepsilon$ -dominates the solutions  $y, z, w$ , i.e., none of the following preferences  $y \succsim_{\varepsilon} x$ ,  $z \succsim_{\varepsilon} x$  and  $w \succsim_{\varepsilon} x$  holds.

More generally, the notion of covering can be introduced as follows:

**Definition 2.** For all  $\varepsilon > 0$  a subset  $B \subseteq A$  of solutions is said to be a  $\varepsilon$ -covering of  $A$  when:  $\forall x \in A, \exists y \in B, y \succsim_{\varepsilon} x$ .

In general, for a given  $\varepsilon$ , several  $\varepsilon$ -coverings exist of different cardinalities, the most interesting being those that are minimal w.r.t. inclusion. The interest of this concept is particularly evident in multicriteria decision problems with a large number of potential solutions. For example, let us consider the multicriteria shortest path problem; in this problem the set of Pareto-optimal solutions can grow exponentially with the number of vertices of the graph. This is well illustrated by the family of bi-criteria graphs introduced in [Hansen, 1980]. In this family, the graph admitting  $2n + 1$  vertices includes  $2^n$  Pareto-optimal solution-paths with distinct cost vectors of the form  $(c, 2^n - 1 - c)$ ,  $c \in \{0, \dots, 2^n - 1\}$ . It then becomes impossible and useless to propose them all to the decision-maker. Instead, we can perform a selection of solutions that  $\varepsilon$ -covers the entire Pareto set. Under the assumption that the criteria values are positive integers and bounded by a quantity  $K$ , Papadimitriou and Yannakakis [Papadimitriou and Yannakakis, 2000] have indeed shown that for any number of criteria  $n \geq 2$  and every set of alternatives  $A$ , there exists a  $\varepsilon$ -covering of  $A$  whose size is bounded from above by  $\lceil \log K / \log(1 + \varepsilon) \rceil^{n-1}$ . This quantity remains polynomial in the size of the problem, when the number of criteria is fixed. For example, in the family of graphs considered by Hansen in [Hansen, 1980] the instance with 33 nodes ( $n = 16$ ) leads to  $2^{16} = 65536$  Pareto-optimal paths with distinct cost vectors of the form  $(c, 2^{16} - 1 - c)$ ,  $c \in \{0, \dots, 2^{16} - 1\}$ . If we choose  $\varepsilon = 0.1$  a covering of the Pareto-optimal cost vectors exists with at most 117 elements; this number decreases to 61 if  $\varepsilon = 0.2$ . For more details on the potential use of the  $\varepsilon$ -dominance and the associated covering concepts see [Papadimitriou and Yannakakis, 2000; Diakonikolas and Yannakakis, 2008; Perny and Spanjaard, 2008; Bazgan et al., 2009].

### Lorenz Dominance

Among the Pareto-optimal solutions, not all are of equal interest to the decision-maker. Some are fairly balanced and show comparable levels of performance on each of the criteria while others alternate excellent performance and very bad points. In multicriteria analysis, the decision-maker often prefers a balanced solution, which does not favor a criterion at the expense of others. A similar principle appears in multi-agent decision problems found with the notion of equity; in this case the criteria measure individual utilities (see chapter I.15). Formally, the idea of equity in the aggregation of preferences can be described by the following axiom, known as the “transfer principle” based on Pigou-Dalton transfers reducing inequalities [Shorrocks, 1983; Moulin, 1988]:

*Transfert principle.* Let  $x \in \mathbb{R}_+^n$  such that  $x_i > x_j$  for  $i, j \in N$ . Then, for any  $\varepsilon$  such that  $0 < \varepsilon \leq x_i - x_j$ ,  $x - \varepsilon e_i + \varepsilon e_j \succ x$  where  $e_i$  (resp.  $e_j$ ) is the vector whose  $i^{\text{th}}$  (resp.  $j^{\text{th}}$ ) component equals 1, all the other components being equal to 0.

This axiom captures the notion of equity as follows: if  $x_i > x_j$ , a mean-preserving shift of performance improving  $x_j$  at the expense of performance  $x_i$  produces a better distribution of criteria satisfaction indices and therefore a better solution. Thus the

vector  $x = (10, 10)$  is preferred to the vector  $y = (14, 6)$  because there is a size 4 transfer from  $y$  to  $x$ . Note that if the transfer was too large, for example  $\varepsilon = 9$ , the inequality would be increased rather than decreased. This is the reason why the  $\varepsilon$  amplitude of the transfer should remain lower than  $x_i - x_j$ .

The transfer principle is a mean-preserving operation since one quantity is removed from one component and added to another. It does not exist between two vectors whose average performances are not the same. However, it becomes more powerful when combined with the Pareto principle requiring that we strictly prefer a solution  $x$  to a solution  $y$  when  $x \succ_P y$ . For example, if one wishes to compare vectors  $(11, 11)$  and  $(12, 9)$ , one can notice that  $(11, 11)$  Pareto-dominates  $(11, 10)$  and that  $(11, 10)$  is deduced from the vector  $(12, 9)$  by a Pigou-Dalton transfer of amplitude 1. Due to the Pareto principle and to the transfer principle, one has on the one hand  $(11, 11) \succ (11, 10)$  and, on the other hand,  $(11, 10) \succ (12, 9)$ , whence, by transitivity,  $(11, 11) \succ (12, 9)$ . The vectors that can be compared by combining the Pareto principle and the transfer principle can be characterized using generalized Lorenz vectors and the notion of generalized Lorenz dominance (see [Marshall and Olkin, 1979; Shorrocks, 1983]):

**Definition 3.** For all  $x \in \mathbb{R}_+^n$ , the *generalized Lorenz vector* associated to  $x$  is defined by:  $L(x) = (x_{\sigma(1)}, x_{\sigma(1)} + x_{\sigma(2)}, \dots, x_{\sigma(1)} + x_{\sigma(2)} + \dots + x_{\sigma(n)})$  where  $\sigma$  represents the permutation which sorts the components of  $x$  by increasing order. Thus  $x_{\sigma(i)}$  represents the  $i^{\text{th}}$  smallest component of  $x$ .

**Definition 4.** The generalized Lorenz dominance is a binary relation defined on  $\mathbb{R}_+^n$  by:  $\forall x, y \in \mathbb{R}_+^n$ ,  $x \succ_L y$  iff  $L(x) \succ_P L(y)$ . The asymmetric part of this relation is defined by  $x \succ_L y$  iff  $L(x) \succ_P L(y)$ .

The  $x$  vector Lorenz-dominates the  $y$  vector if  $L(x)$  Pareto-dominates  $L(y)$ . To show that  $\succ_L$  is indeed an instance of the CA approach, it is sufficient to consider that:

$$\phi_i(x, y) = \begin{cases} 1 & \text{if } \sum_{j=1}^i x_{\sigma(j)} \geq \sum_{j=1}^i y_{\sigma(j)} \\ 0 & \text{otherwise} \end{cases}$$

and  $\psi(z_1, \dots, z_n) = \sum_{i=1}^n z_i - n$ . The notion of Lorenz dominance was initially introduced to compare vectors with the same average (e.g., for comparing various income distributions over a population). The generalized version introduced above is more adapted to the context of multicriteria optimization because it makes it possible to compare vectors of performances that do not have the same average. The link between generalized Lorenz dominance and the transfer principle appears with the following result [Chong76]:

**Theorem 1.** For all pairs of vectors  $x, y \in \mathbb{R}_+^n$ , if  $x \succ_P y$ , or if  $x$  is obtained from  $y$  using a Pigou-Dalton transfer, the  $x \succ_L y$ . Conversely, if  $x \succ_L y$ , then there exists a sequence of Pigou-Dalton transfers and/or Pareto improvements allowing to pass from  $y$  to  $x$ .

This result establishes the generalized Lorenz dominance as the minimal relation w.r.t. inclusion which simultaneously satisfies the Pareto principle and the principle

of transfer. This dominance is transitive. To illustrate the use of Lorenz dominance to compare the vectors considered above, one can observe that  $L(11, 10) = (10, 21)$  while  $L(12, 9) = (9, 21)$ . We thus have  $(11, 10) \succ_L (12, 9)$  since  $(10, 21) \succ_P (9, 21)$ .

A consequence of the preceding theorem is that if  $x \succ_P y$  then  $x \succ_L y$ , which shows that the Lorenz dominance is potentially more discriminating than the Pareto dominance. It follows that  $ND(\mathcal{X}, \succ_L) \subseteq ND(\mathcal{X}, \succ_P)$ , that is to say that the non-dominated solutions in Lorenz's sense are Pareto-optimal solutions. Apart from a few specific cases, there are generally significantly fewer Lorenz-optimal than Pareto-optimal solutions. The Lorenz dominance thus appears as a natural refinement of the Pareto dominance allowing to remove unfair elements in the Pareto set.

### Weighted Lorenz Dominance

The Lorenz dominance deals symmetrically with all components of the vectors that are compared. The  $L(x)$  vector indeed remains invariant by permutation of the components of  $x$  and consequently the preference  $x \succ_L y$  is not affected by a permutation of the components of  $x$  or  $y$ . This characteristic seems natural when one wishes to assign the same importance to all criteria or agents. On the other hand, if we want to give more weight to some of the criteria, we should consider a weighted extension of the Lorenz dominance. A first idea that naturally comes to mind to assign different weights to components is to duplicate them in proportion to the weights of the criteria (we assume here that the weights are rational numbers). Thus, if we want to compare the vectors  $x = (10, 5, 15)$  and  $y = (10, 12, 8)$  given that the criteria have weights defined by the vector  $p = (3/6, 1/6, 2/6)$  we can consider the extended vectors  $\tilde{x} = (10, 10, 10, 5, 15, 15)$  and  $\tilde{y} = (10, 10, 10, 12, 8, 8)$  and test if  $\tilde{x} \succ_L \tilde{y}$  or  $\tilde{y} \succ_L \tilde{x}$ . This is not the case here since the Lorenz vectors  $(5, 15, 25, 35, 50, 65)$  and  $(8, 16, 26, 36, 46, 68)$  are incomparable in terms of Pareto dominance. Here, the fact that criterion 3 is twice more important than criterion 1 did not allow us to prefer  $y$  although  $y$  distributes performance more equally than  $x$ .

A more elaborate way of proposing a weighted extension of the Lorenz dominance without having to duplicate the components (nor assuming that weights are rational numbers) is to associate to each vector  $x$  a cumulative function  $F_x(z)$  which indicates the weight of the coalition formed by the criteria whose performance does not exceed threshold  $z$ . Denoting  $v$  the function which gives the weight of a subset of criteria, we have:  $F_x(z) = v(\{i \in N, x_i \leq z\})$ . We also consider the left inverse of  $F_x$ , denoted  $\check{F}_x$  which reads  $\check{F}_x(p) = \inf\{z \in \mathbb{R} | F_x(z) \geq p\}$  for  $p \in [0, 1]$ . This quantity represents a kind of quantile function; it represents the minimum level  $z$  from which there exists a coalition of criteria satisfied at level  $z$  or more and which is of weight greater than or equal to  $p$ . Both  $F_x$  and  $\check{F}_x$  are stepwise functions. We then define from  $F_x, F_y$  or  $\check{F}_x, \check{F}_y$  the second order stochastic dominance by one of the following formulas which are known to be equivalent:

$$x \succsim_2 y \text{ iff } \forall z \in \mathbb{R}, F_x^2(z) \leq F_y^2(z) \quad \text{with } F_x^2(z) = \int_{-\infty}^z F_x(t) dt \quad (7)$$

$$x \succsim_2 y \text{ iff } \forall p \in [0, 1], \check{F}_x^2(p) \geq \check{F}_y^2(p) \text{ with } \check{F}_x^2(p) = \int_0^p \check{F}_x(t) dt \quad (8)$$

This dominance is transitive and coincides with the second order stochastic dominance that will be introduced in the next chapter (just reinterpreting the  $v$  function as a probability measure and using Equation (7)). In the case where the criteria are equally weighted, this dominance  $\succsim_2$  reduces to the Lorenz dominance. In fact, when functions  $F_x, F_y$  (resp.  $\check{F}_x, \check{F}_y$ ) are piecewise linear, testing  $\succsim_2$  amounts to comparing the curves at break points in terms of weak Pareto dominance. Note that in the case of equally weighted components, the  $n$  break points are in  $k/n$  for  $k = 1, \dots, n$ . We can then show that  $n\check{F}_x^2(k/n) = L_k(x)$  (see [Shorrocks, 1983; Muliere and Scarsini, 1989]). This explains why  $\succsim_2$  is equivalent to comparing the components of the Lorenz vectors  $L_k(x)$  and  $L_k(y)$  for all  $k \in N$  and therefore, in this case,  $\succsim_2$  is nothing else but the Lorenz dominance  $\succsim_L$ . For this reason  $\succsim_2$  can be seen as a generalization of Lorenz dominance in the case of weighted criteria.

### 3.2 Concordance Relations

The concordance relations are preference relations that are not necessarily transitive, resulting from aggregation rules inspired by the majority voting rules (rules of concordance). In such rules, for each pair of solutions  $(x, y)$ , we count the number of criteria in favor of  $x$  and  $y$  respectively, and it is based on this count to decide whether  $x$  is better than  $y$ . If the criteria do not all have the same weight, we can more generally evaluate the weight of the coalition of criteria in favor of  $x$  and against  $y$ . This so-called ‘‘concordant’’ coalition is a widely used notion in ELECTRE type methods (see [Roy, 1985; Roy and Bouyssou, 1993; Vincke, 1992]). There are many variants of these rules, of which we give here some typical examples by assuming that the indices  $\phi_i(x, y)$  are constructed as in Equation (4):

#### Absolute Concordance

$$\forall (x, y) \in A \times A, c(x, y) = v(\{i \in N : \phi_i(x, y) > 0\}) \quad (9)$$

$$x \succsim y \text{ iff } c(x, y) \geq s \quad (10)$$

where  $v$  is a set function defined on  $2^N$  and valued in  $[0, 1]$  and  $s \in ]0, 1]$  is an acceptance threshold named *concordance threshold*. The most standard instance of this family of rules is the absolute majority rule obtained for  $s = (n + 1)/2$  and  $v(E) = |E|$  for all  $E \subseteq N$ . When we wish to weight the criteria, we can define  $v(E) = \sum_{i \in E} w_i$  where  $w_i$  represents the weight of criterion  $i$ . We may also resort to more general definitions such as  $v(E) = \Psi(w_1, \dots, w_n)$  where  $\Psi$  is an aggregation function.

### Relative Concordance

$$x \succsim y \text{ iff } c(x, y) \geq c(y, x) \quad (11)$$

where  $c(x, y)$  is defined by Equation (9). This relation is generally not transitive. However, some instances of this rule are transitive on any set of alternatives. In fact, the transitivity appears for very specific definitions of the notion of importance of criteria. This point has been widely studied in the literature on decision theory, see e.g., [Dubois et al., 2003] for a reference in AI.

### Absolute Concordance with Veto

$$x \succsim y \text{ iff } \begin{cases} c(x, y) \geq s \\ \forall i \in N, y_i - x_i \leq v_i \end{cases} \quad (12)$$

where  $v_i$  is the veto threshold that can be defined as the biggest difference of performance  $y_i - x_i$  that can be imagined on criterion  $i$  and which is still compatible with the preference of  $x$  over  $y$ . If  $y_i - x_i$  exceeds the veto threshold on some criterion  $i$ ,  $x$  cannot be preferred nor be indifferent to  $y$ . This condition aims to prevent any compensation phenomenon when comparing two alternatives with very contrasted profiles. It also prevents to compensate a strong weakness with multiple weakly positive points. This principle of non-veto is presented in an absolute concordance rule but could also be inserted in the relative concordance rule. The reader may refer to [Roy and Bouyssou, 1993; Perny, 1998] for more details on this point.

### Concordance Rules with Reference Points

Let  $p \in \mathbb{R}^n$  be a performance vector used as a reference point to assess and compare the alternatives. A concordance relation with reference point is defined by:

$$x \succsim y \text{ iff } c(x, p) \geq c(y, p) \quad (13)$$

where  $c(x, y)$  is defined by Equation (9). Using the same notations, we can also introduce the following relation:

$$x \succsim y \text{ iff } c(p, y) \geq c(p, x). \quad (14)$$

Note that, contrary to the standard concordance relations introduced before (see Equations 10 to 12), the concordance relations with reference point are naturally transitive, which facilitates their use for choice and ranking problems. One can find in [Perny and Rolland, 2006; Rolland, 2013; Bouyssou and Marchant, 2013] other examples of concordance rules with reference points, as well as some axiomatic analysis concerning these rules.

When a non-transitive concordance relation is used, the candidates cannot be directly ordered and it is difficult to determine an optimal choice. To overcome the

problem, many methods for determining a winner or ranking the alternatives from a non-transitive strict preference relation  $\succ$  have been proposed. Here are some examples:

### The Net Flow Rule

Rank the candidates according to the Net Flow Score defined as follows:

$$\phi(x) = |\{y \in \mathcal{X} : x \succ y\}| - |\{y \in \mathcal{X} : y \succ x\}|$$

For a choice problem, select the alternatives maximizing the net flow.

### Schwartz's Rule

Calculate  $\succ^*$  the transitive closure of relation  $\succ$ . Then define a new strict preference relation  $\succ_S$  as follows:

$$x \succ_S y \text{ iff } [x \succ^* y \text{ and not}(y \succ^* x)]$$

By construction relation  $\succ_S$  is transitive since it is the asymmetric part of a transitive relation. For a choice problem, select the solutions of  $ND(\mathcal{X}, \succ_S)$ .

### Decision Rules based on Traces

The traces of a relation  $\succ$  are defined by:

$$\begin{aligned} x \succ^+ y &\text{ iff } \forall z \in \mathcal{X} \setminus \{x, y\}, (y \succ z \Rightarrow x \succ z) \\ x \succ^- y &\text{ iff } \forall z \in \mathcal{X} \setminus \{x, y\}, (z \succ x \Rightarrow z \succ y) \end{aligned}$$

Both relations  $\succ^+$  and  $\succ^-$  are transitive, and therefore their intersection too. They can therefore be used to partially order the solutions or to define a set of non-dominated elements, for example by calculating  $ND(\mathcal{X}, \succ^+)$  or  $ND(\mathcal{X}, \succ^-)$ .

## 4 Decision Models in the AC Approach

### 4.1 The weighted Mean

The decision model based on the weighted mean leads to the following definition of preferences:

$$x \succsim y \text{ iff } \sum_{i=1}^n w_i x_i \geq \sum_{i=1}^n w_i y_i$$

This model is probably the one that most quickly comes to mind when one aggregates performances. Yet it is often unsatisfactory because it provides no control on whether the optimal solutions are balanced or not. By way of illustration, let us consider the following example:

*Example 5.* A company wants to recruit a technical sales computer. Candidates must complete two interviews, one for the technical skills of the individual, the other intended to evaluate the business skills. Consider a situation with 4 candidates that received the following grades:  $x = (18, 5)$ ,  $y = (4, 19)$ ,  $z = (11, 11)$ ,  $w = (9, 7)$ . Candidate  $w$  who is Pareto-dominated by candidate  $z$  is quickly disqualified. The candidates  $x$  and  $y$ , who have a significant weakness on one of the two expected skills (score less than or equal to 5), do not seem to be suitable either. As a result, the  $z$  candidate seems to be the best compromise between technical and commercial skills. However, it can easily be verified that, whatever the weight vector  $(w_1, w_2)$  used, the candidate  $z$  will not be the one with the best weighted average, although he is Pareto-optimal. This is due to the fact that  $(11, 11)$  lies within the convex hull of the points  $x, y, z, w$  in the criterion space, whereas only the points on the boundaries of this convex hull can be obtained by optimizing a weighted sum of the performances.

The above example shows that when a weighted sum is used, we take the risk of eliminating some Pareto-solutions *a priori*, even before having chosen the weights of the criteria, although such solutions could achieve interesting compromises between the criteria. These well-known limits of the weighted sum justify the interest in other aggregators. A possible generalization of weighted means is provided by quasi-arithmetic means defined by:

$$\psi(x) = f^{-1} \left( \sum_{i=1}^n w_i f(x_i) \right)$$

where  $f(x)$  is a strictly monotonic function. For instance, the weighted geometric mean is obtained for  $f(x) = \ln(x)$ , the weighted dual geometric mean when  $f(x) = \ln(1-x)$ , the harmonic mean when  $f(x) = 1/x$  and the weighted  $L_p$  norm for  $f(x) = x^p$ ,  $p \in \mathbb{N}$ . The next section introduces a more powerful aggregator to explore various types of compromise solutions in the Pareto set.

## 4.2 The Weighted Tchebycheff Norm

One way to define preferences by a scalarizing function is to measure the distance to a reference point  $p \in \mathbb{R}^n$  representing a target performance vector. The idea is to try to be as close as possible to the target on each of the criteria. The quality of a solution can then be defined as its distance to the target in the sense of the Tchebycheff norm (a.k.a. infinite norm).

Let  $\lambda \in \mathbb{R}_+^n$  be a weighting vector used in combination with the Tchebycheff norm on the one hand to normalize criterion values when they are expressed on different scales, and on the other hand, to control the importance attached to the different criteria so as to generate compromise solutions with a bias reflecting the value system of the decision maker:

$$\psi(x) = \|\lambda(x - p)\|_\infty = \max_{i \in N} \lambda_i |x_i - p_i|$$

A good choice for the reference point  $p$  is the *ideal point*  $\alpha \in \mathbb{R}^n$  defined by  $\alpha_i = \sup_{x \in \mathcal{X}} x_i$  which provides an upper bound of the set of Pareto-optimal vectors. Then, minimizing the Tchebycheff distance amounts to projecting the ideal point on the Pareto set, in a direction controlled by the weights  $\lambda_i$ . Usually the weights  $\lambda_i$  are defined as follows:

$$\lambda_i = \frac{w_i}{\alpha_i - \beta_i} \text{ with } \beta_i = \inf_{x \in \mathcal{X}^*} \{x_i\} \text{ and } \mathcal{X}^* = \{x \in \mathcal{X} : \exists i \in N, x_i = \alpha_i\}$$

The components  $\alpha_i$  are obtained by monocriterion optimizations on each component separately; this makes it possible to compute  $\mathcal{X}^*$  and then components  $\beta_i$ . The optimization of the parameterized function  $\psi$  guarantees that, for any Pareto-optimal solution  $x$ , there exists a weighting vector  $w$  such that  $x$  will be part of the  $\psi$ -optimal solutions [Wierzbicki, 1986] (in fact to avoid pathological cases and fully benefit from this property, it is better not to define  $\alpha$  as the ideal point but as a neighbor point strictly above the ideal on every component). We thus correct the observed defect of the weighted sum since any Pareto-optimal solution can now be obtained by a minimization of  $\psi$  with the proper parameters. On the other hand, the optimization of function  $\psi$  does not quite guarantee the Pareto-optimality of the solutions obtained because of a drowning effect induced by the maximum. Assume indeed that the reference point is  $p = (20, 20)$  and that the two feasible solutions are  $x = (4, 2)$  and  $y = (18, 2)$  we have  $\psi(x) = \psi(y)$ . Thus,  $x$  could be selected as the best choice while it is Pareto-dominated. To avoid this problem, we introduce an additional term, the weighted sum of the deviations from the ideal point multiplied by an arbitrarily small quantity  $\varepsilon > 0$ ; this weighted sum comes to play the role of a second criterion considered lexicographically after that of Tchebycheff to discriminate between equivalent solutions in terms of distance to the ideal point. We therefore arrive at the following aggregation function to be minimized:

$$t(x) = \max_{i \in N} \frac{w_i(\alpha_i - x_i)}{\alpha_i - \beta_i} + \varepsilon \sum_{i=1}^n \frac{w_i(\alpha_i - x_i)}{\alpha_i - \beta_i} \quad (15)$$

By minimizing function  $t$  defined by Equation (15), we make sure to generate only Pareto-optimal solutions. Moreover, if  $\varepsilon$  is chosen to be small enough, the practical possibility of reaching any Pareto-optimal solution is preserved [Wierzbicki, 1986]. This dual quality justifies the use of this aggregator in optimization to explore the Pareto-optimal solutions in various directions controlled by the  $w$  vector. It is therefore widely used in interactive exploration methods [Steuer and Choo, 1983;

Steuer, 1986; Wierzbicki, 1999]. This aggregator can of course be used to define a preference over the set of solutions (by proximity to the ideal point  $\alpha$ ) by setting:  $x \succsim y$  iff  $t(x) \leq t(y)$ . An application of this decision model to multiobjective state space search is proposed in [Galand and Perny, 2006].

### 4.3 The Ordered Weighted Average (OWA)

The Ordered Weighted Average [Yager, 1988] is an aggregation function enabling to weight the performances  $x_i$  according to their rank, once reordered with permutation  $\sigma$  such that  $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}$ . Formally, the OWA aggregation is defined by:

$$OWA(x) = \sum_{i=1}^n w_i x_{\sigma(i)} \quad (16)$$

OWA is a symmetric function because the weights do not relate to the components of  $x$  but to those of the reordered vector. This family includes the minimum, maximum, median operations and all order statistics<sup>2</sup> as special cases, by using a weighting vector whose all but one components are zero, the remaining one being 1. It is also widely used in fair optimization as a linear extension of the Lorenz dominance introduced in the previous section. Indeed, noting that  $x_{\sigma(i)} = L_i(x) - L_{i-1}(x)$  for  $i > 1$ , we have:

$$OWA(x) = \sum_{i=1}^{n-1} (w_i - w_{i+1}) L_i(x) + w_n L_n(x) \quad (17)$$

One can see that, if coefficients  $w_i$  are positive and chosen to decrease when  $i$  increases, OWA is a linear combination with positive coefficients of the components of the Lorenz vector. Therefore the weak order defined by:

$$x \succsim y \text{ iff } OWA(x) \geq OWA(y)$$

is a linear extension of Lorenz dominance, that is,  $x \succsim_L y \Rightarrow OWA(x) \geq OWA(y)$ . Thus OWA used with strictly decreasing weights  $w_i$  is an aggregator allowing to promote balanced solutions. Indeed, due to Equation (17), an OWA-optimal solution is necessarily optimal in Lorenz's sense and there is therefore no Pigou-Dalton transfer allowing to reduce inequalities (due to Theorem 1). Another way to present the treatment of inequalities by an OWA is to consider Equation (16) and note that by choosing decreasing weights, one assigns the greatest weight to the least satisfied criterion, then a little less importance to the second least satisfied criterion and so on. Of course, comparing or sorting performances from several criteria only makes sense when they are expressed on the same scale (if not, they must first be re-encoded using utility functions). To give an example of the use of OWA, if one wishes to compare the vectors  $x = (10, 5, 15)$  and  $y = (10, 12, 8)$  us-

<sup>2</sup> The order statistic of rank  $k$  of a sample of values is equal to the  $k^{th}$  smallest value.

ing an OWA with  $w = (3/6, 2/6, 1/6)$ , we obtain  $OWA(x) = 50/6 = 8.33$  while  $OWA(y) = 52/6 = 8.66$ , therefore  $y \succ x$ .

The OWA operator is widely used in social choice theory as a measure of inequality under the name of the generalized Gini social evaluation function [Weymark, 1981]. It is also used to aggregate fuzzy set membership functions (see [Yager, 1988]). In artificial intelligence, it often appears in fair optimization problems or allocation problems of indivisible goods [Bouveret and Lang, 2005; Golden and Perny, 2010; Lesca and Perny, 2010], and also in voting rules [Goldsmith et al., 2014; Elkind and Ismaili, 2015; Skowron et al., 2016; García-Lapresta and Martínez-Panero, 2017]. Note that, although OWA is not a linear function of criterion values, the optimization of an OWA function can be done by linear programming (provided that the criteria and the constraints defining the admissible solutions are linear in the decision variables), for more details see [Ogryczak and Sliwinski, 2003; Chassein and Goerigk, 2015].

#### 4.4 The Weighted OWA (WOWA)

As pointed out in the previous subsection, one characteristic of the OWA is to be a symmetric aggregation function. This property, which seems natural when the criteria represent individual points of view in a collective decision problem, may not be desired in multicriteria decision problems, particularly when certain criteria are considered more important than others. We then consider now a weighted extension of the OWA aggregator, the initial weights involved in the OWA definition being only used to control the importance attached to good and bad performances. This weighted OWA is known in the literature under the name of WOWA [Torra, 1997]; it uses a vector  $p \in \mathbb{R}^n$  of criteria weights and takes the following form:

$$\begin{aligned} WOWA(x) &= \sum_{i=1}^n [x_{\sigma(i)} - x_{\sigma(i-1)}] \varphi \left( \sum_{k=i}^n p_{\sigma(k)} \right) \\ &= \sum_{i=1}^n \left[ \varphi \left( \sum_{k=i}^n p_{\sigma(k)} \right) - \varphi \left( \sum_{k=i+1}^n p_{\sigma(k)} \right) \right] x_{\sigma(i)} \end{aligned}$$

where  $\sigma$  is the permutation reordering the components of  $x$  by increasing order, i.e.,  $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}$ ; function  $\varphi$  is strictly increasing and such that  $\varphi(0) = 0$ . The induced preference is then defined by:  $x \succsim y$  iff  $WOWA(x) \geq WOWA(y)$ .

This formulation is known as *Yaari's model* in decision under risk because it has been initially introduced and axiomatically justified in this context [Yaari, 1987] (the weights  $p_i$  being interpreted as the probabilities of the states of nature, see the RDU model in Chapter I.14). Its importation into a multicriteria decision-making context is more recent and due to Torra [Torra, 1997] who arrives at an identical formulation starting from an OWA. The specificity of the construction proposed by Torra lies in

the definition of the  $\varphi$  function. It is constructed from the weights  $(w_1, \dots, w_n)$  of an OWA so that  $\varphi(i/N) = \sum_{k=1}^i w_{n-k+1}$ , which allows to weight the criteria while controlling the importance given to good and bad performances as in an OWA. The WOWA is therefore constructed from the two weighting vectors  $p$  and  $w$ . By way of illustration, consider the following example:

*Example 6.* If we want to compare vectors  $x = (10, 5, 15)$  and  $y = (10, 12, 8)$  with a WOWA characterized by weights  $w = (3/6, 2/6, 1/6)$  and criteria weights  $p = (3/6, 1/6, 2/6)$ , we use a piecewise linear  $\varphi$  function taking the following values at the key points:  $\varphi(0) = 0$ ,  $\varphi(1/3) = 1/6$ ,  $\varphi(2/3) = 1/2$ ,  $\varphi(1) = 1$ . These values can be completed by linear interpolation to obtain:

$x$	0	1/6	2/6	3/6	4/6	5/6	1
$\varphi(x)$	0	1/12	1/6	2/6	1/2	3/4	1

Then we get:

$$WOWA(x) = 5 + (10 - 5)\varphi(5/6) + (15 - 10)\varphi(2/6) = 9.58$$

$$WOWA(y) = 8 + (10 - 8)\varphi(4/6) + (12 - 10)\varphi(1/6) = 9.16$$

Therefore  $x$  is preferred to  $y$ . Here the fact that the third criterion is more important than the second gives an advantage to  $x$  which is sufficient to compensate the inegalitarian side of this solution. We can verify that it would suffice to be more demanding on the equity requirement by choosing the weighting vector  $w = (0.8, 0.25, 0.05)$  so that the preference is reversed in favor of  $y$ . In this case, we would indeed have:

$$\varphi(0) = 0, \varphi(1/3) = 0.05, \varphi(2/3) = 0.3, \varphi(1) = 1$$

and, by completing using linear interpolation:

$x$	0	1/6	2/6	3/6	4/6	5/6	1
$\varphi(x)$	0	0.025	0.05	0.175	0.3	0.65	1

In this case we obtain:

$$WOWA(x) = 5 + (10 - 5)\varphi(5/6) + (15 - 10)\varphi(2/6) = 8.5$$

$$WOWA(y) = 8 + (10 - 8)\varphi(4/6) + (12 - 10)\varphi(1/6) = 8.65$$

and this time we get  $y$  is preferred to  $x$ .

This example clearly shows how the two vectors of weights interact, one to control the weights of the criteria and the other to control the fairness requirement. It may also be noted that if weights  $w_i$  decrease when  $i$  increases (to favor solutions that equitably share performance among criteria) then the  $\varphi$  function is convex and the function WOWA is concave. In this case, it can be proven that WOWA is monotone increasing with dominance  $\succsim_2$  introduced in Equation (7), which means that:

$x \succsim_2 y \Rightarrow WOVA(x) \geq WOVA(y)$ . Thus, we obtain a weighted version of the result concerning the monotony of OWA with respect to Lorenz dominance.

In practice, nothing prohibits the use of non-decreasing weights and we then obtain an aggregator that offers a diversity of behaviors in the aggregation. For example, if we use weights  $w_i$  increasing with  $i$ , we will have a concave  $\varphi$  and a convex WOWA. In maximization, we give a premium to the solutions with an imbalanced profile alternating good and bad performances. We will return to the control of WOWA in the more general framework of the Choquet integral. We can also notice that if  $\phi(x) = x$  then WOWA reduces to a simple weighted average with weights  $p_i$ . Moreover, if  $p_i = 1/n$  then WOWA reduces to an OWA, which is quite natural. Finally, it should be noted that WOWAs are very useful for equitable optimization when we want to associate weights with agents (exogenous rights), especially since when the  $w_i$  weights are decreasing when  $i$  increases, the optimization of WOWA can be done easily by linear programming using reformulations close to those needed to linearize an OWA (see [Ogryczak and Sliwinski, 2007] for more details).

Note that function  $\varphi$  does not necessarily have to be constructed from vectors  $p$  and  $w$ , it can be directly defined as a convex function to convey an idea of fairness (in maximization, the more function  $\varphi$  is convex the greater the requirement of fairness). On the contrary a concave  $\varphi$  function would exhibit a preference for contrasted profiles. In minimization problems, this is just the opposite and fairness is modeled by a concave  $\varphi$  function. The direct elicitation of  $\varphi$  can be performed using preference queries on specific pairs of alternatives, see e.g., [Perny et al., 2016].

#### 4.5 The Choquet Integral

The Choquet integral is one of the most sophisticated scalarizing function used for multicriteria aggregation [Choquet, 1953; Grabisch, 1996; Marichal, 2000a; Marichal and Roubens, 2000; Grabisch and Labreuche, 2008; Grabisch et al., 2009]. It includes both weighted sums, OWA and WOWA as special cases. It is defined from a set function, namely the *capacity* allowing to assign a weight to any subset of criteria  $E \subseteq N$ . More precisely, a capacity is defined as follows:

**Definition 5.** A capacity is a set function  $v : N \rightarrow [0, 1]$  such that  $v(\emptyset) = 0$ ,  $v(N) = 1$  and  $\forall A, B \subseteq N, A \subseteq B \Rightarrow v(A) \leq v(B)$ .

The capacity is said to be:

- *concave* or *sub-modular* if  $v(A \cup B) + v(A \cap B) \leq v(A) + v(B)$  for all  $A, B \subseteq N$ ,
- *additive* if  $v(A \cup B) + v(A \cap B) = v(A) + v(B)$  for all  $A, B \subseteq N$ ,
- *convex* or *super-modular* if  $v(A \cup B) + v(A \cap B) \geq v(A) + v(B)$  for all  $A, B \subseteq N$ .

For any given capacity  $v$ , the Choquet integral of a vector  $x \in \mathbb{R}^n$  is defined by:

$$\begin{aligned}
C_v(x) &= \sum_{i=1}^n [x_{\sigma(i)} - x_{\sigma(i-1)}] v(X_{\sigma(i)}) \\
&= \sum_{i=1}^n [v(X_{\sigma(i)}) - v(X_{\sigma(i+1)})] x_{\sigma(i)}
\end{aligned}$$

where  $\sigma$  is the permutation reordering the components of  $x$  by increasing order, i.e.,  $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}$ ,  $x_{\sigma(0)} = 0$ ,  $X_{\sigma(i)} = \{\sigma(i), \sigma(i+1), \dots, \sigma(n)\}$  for  $i = 1, \dots, n$ , and  $X_{\sigma(n+1)} = \emptyset$ . The preference relation associated to the Choquet integral is therefore defined by:  $x \succsim y$  iff  $C_v(x) \geq C_v(y)$ .

We remark that WOWA is only a special case of Choquet integral in which the capacity  $v$  is defined by  $v(E) = \varphi(\sum_{i \in E} P_i)$  for a weighting vector  $(p_1, \dots, p_n)$ . It can be shown that  $v$  is convex (resp. concave) when  $\varphi$  is convex (resp. concave). The Choquet integral  $C_v$  can account for various behaviors depending on the choice of the capacity. When using an additive capacity, i.e.,  $v(E) = \sum_{i \in E} p_i$ , the Choquet integral is reduced to the weighted sum and does not offer particular descriptive possibilities. However, we can describe much richer classes of preferences with concave, convex or other more general capacities. For example, if we use a convex capacity in maximization, the well-balanced profiles will be favored, and this will be the opposite if we choose a concave capacity, as shown by the following example:

*Example 7.* Let us consider an example with 3 criteria, i.e.,  $N = 1, 2, 3$  and two capacities  $v_1$  and  $v_2$  defined in the following table:

	$\emptyset$	$\{1\}$	$\{2\}$	$\{3\}$	$\{1, 2\}$	$\{1, 3\}$	$\{2, 3\}$	$\{1, 2, 3\}$
$v_1$	0	0.2	0.1	0.3	0.45	0.5	0.65	1
$v_2$	0	0.35	0.5	0.55	0.7	0.9	0.8	1

One can easily check that  $v_1$  is convex and  $v_2$  is concave. Coming back to the comparison of vectors  $x = (10, 5, 15)$  and  $y = (10, 12, 8)$  we have:

$$\begin{aligned}
C_{v_1}(x) &= 5 + (10 - 5)v_1(\{1, 3\}) + (15 - 10)v_1(\{3\}) = 9 \\
C_{v_1}(y) &= 8 + (10 - 8)v_1(\{1, 2\}) + (12 - 10)v_1(\{2\}) = 9.1 \\
C_{v_2}(x) &= 5 + (10 - 5)v_2(\{1, 3\}) + (15 - 10)v_2(\{3\}) = 12.25 \\
C_{v_2}(y) &= 8 + (10 - 8)v_2(\{1, 2\}) + (12 - 10)v_2(\{2\}) = 10.4
\end{aligned}$$

One can see that with  $v_1$  solution  $y$  is preferred to  $x$  whereas with  $v_2$  solution  $x$  is preferred to  $y$ .

More precisely, the use of a convex capacity in a Choquet integral conveys an idea of equity due to the following property [Chateauneuf and Tallon, 1999]:

**Proposition 1.** *If  $v$  is convex then  $\forall x^1, x^2, \dots, x^p \in \mathbb{R}^n$ ,  $\forall k \in \{1, 2, \dots, p\}$  and  $\forall i \in \{1, 2, \dots, p\}$ ,  $\lambda_i > 0$  such that  $\sum_{i=1}^p \lambda_i = 1$  we have:*

$$C_v(x^1) = C_v(x^2) = \dots = C_v(x^p) \Rightarrow C_v\left(\sum_{i=1}^p \lambda_i x^i\right) \geq C_v(x^k).$$

This proposition means that if several solutions  $x^i, i = 1, \dots, p$  are indifferent for the decision maker, she will prefer a solution whose performance vector is a convex combination of the  $x^i$ 's to all these solutions. For example, if one is indifferent between two performance vectors  $(0, 20)$  and  $(20, 0)$ , it is expected that a solution such as  $(10, 10)$  which corresponds to the average of the two preceding vectors is preferable. Obviously, the reverse preference is obtained with a concave capacity. The Choquet integral for a convex capacity is a concave function and conversely the Choquet integral for a concave capacity is a convex function [Lovász, 1983]. Another useful formulation of the Choquet integral is to express it as a function of the Möbius masses associated with capacity  $\nu$ . These masses are defined in this way:

**Definition 6.** To any capacity  $\nu$  defined on  $2^N$  one can associate another set function on  $2^N$  named *Möbius inverse* and defined by:

$$\forall A \subseteq N, m(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} \nu(B) \quad (18)$$

Then,  $\nu$  can be recovered from its Möbius inverse  $m$  as follows:

$$\forall A \subseteq N, \nu(A) = \sum_{B \subseteq A} m(B) \quad (19)$$

Using the Möbius inverse  $m$  associated with  $\nu$ , the Choquet integral can be rewritten as follows:

$$C_\nu(x) = \sum_{B \subseteq N} m(B) \min_{i \in B} x_i \quad (20)$$

This highlights another interpretation of the Choquet integral as a linear aggregator in a new multidimensional space of size  $2^n$  where  $n$  is the initial number of criteria. The components of a vector in this space correspond to quantities  $\min_{i \in B} x_i$  for all subsets  $B \subseteq N$ . Whether we use the initial formulation of the Choquet integral or the one that involves the Möbius masses, we may be concerned about the presence of  $2^n$  parameters to characterize the importance of the criteria and their interaction. Fortunately, in many practical cases, there is no need to consider all these coefficients, we can resort to *k-additive* capacities for some  $k < n$ , where *k-additivity* is defined as follows [Grabisch, 1996; Dubois and Prade, 1997]:

**Definition 7.** A capacity  $\nu$  is said to be *k-additive* if its Möbius inverse equals zero for any subset  $A \subseteq N$  such that  $|A| > k$ , and if  $m(A) \neq 0$  for some  $A$  such that  $|A| = k$ .

If  $k = 1$  we obtain an additive capacity. The *k-additive* capacities for small values of  $k$  greater than 1 are very useful in practice because they offer sufficient expressiveness to model positive or negative interactions between criteria while involving a fairly small number of parameters. For example, when  $k = 2$ , the capacity is completely characterized by only  $(n^2 + n)/2$  coefficients (a Möbius mass for each singleton and each pair), which enables to model the following interactions between pairs of criteria:

- positive interaction:  $m(\{i, j\}) > 0$  and therefore  $\nu(\{i, j\}) > \nu(\{i\}) + \nu(\{j\})$

- no interaction:  $m(\{i, j\}) = 0$  and therefore  $v(\{i, j\}) = v(\{i\}) + v(\{j\})$
- negative interaction:  $m(\{i, j\}) < 0$  and therefore  $v(\{i, j\}) < v(\{i\}) + v(\{j\})$

Moreover, with a 2-additive capacity one obtains from Equation (20) a very compact expression for the Choquet integral:

$$C_v(x) = \sum_i m_i x_i + \sum_{i>j} m_{ij} \min\{x_i, x_j\}$$

As with OWA and WOWA, the search for a solution maximizing  $C_v(x)$  can be performed by linear programming in the case where  $v$  is convex [Lesca and Perny, 2010]. In the general case, it is more delicate but some efficient linearizations exist for some class of Möbius representations [Lesca et al., 2013]. For more details on Choquet integrals, interaction indices and set functions in multicriteria analysis, the reader should refer to [Grabisch, 1996; Grabisch et al., 2009; Grabisch, 2016]. For the elicitation of the capacity in the Choquet integral, the reader should refer to [Grabisch et al., 1995; Marichal and Roubens, 2000; Fallah Tehrani et al., 2012; Hüllermeier and Fallah Tehrani, 2013; Benabbou et al., 2017].

The Choquet integral is used in various domains of artificial intelligence. For example, in machine learning, the use of Choquet integrals provides higher predictive capacities than linear models, while offering measures for quantifying the importance of individual predictor variables and the interaction between groups of variables [Fallah Tehrani et al., 2012]. Moreover, in recommender systems [Beliakov et al., 2015], the advantage provided by Choquet integrals is to allow positive and negative synergies between criteria, with enhanced descriptive and prescriptive possibilities. Similarly, in multiagent decision making [Dubus et al., 2009], the Choquet integral is used to aggregate individual preferences using a possibly non-additive measure of the importance of agent coalitions, which allows one to model various notions of social welfare. In information fusion [Torra and Narukawa, 2007], the use of the Choquet integral allows one to model positive or negative reinforcements among sets of observations. Finally, in multiobjective state-space search [Galand and Perny, 2007], the use of Choquet integrals allows one to find compromise solutions that could not be obtained using linear aggregators.

#### 4.6 The Sugeno Integral

The Sugeno integral [Sugeno, 1974; Dubois et al., 1998; Marichal, 2000b; Dubois et al., 2001a; Grabisch and Labreuche, 2008; Couceiro et al., 2012] can be seen as a qualitative counterpart of the Choquet integral. In some cases, performance and capacity are expressed on a common ordinal scale. In the presence of such information, one cannot reasonably use the previous criteria which call for the cardinal properties of performance and importance indices (weight, capacity). A natural alternative is then to consider the Sugeno integral which reads:

$$S_v(x) = \max_{i \in N} \min\{x_{\sigma(i)}, v(X_{\sigma(i)})\}$$

where  $\sigma$  is the permutation reordering the components of  $x$  by increasing order, i.e.,  $x_{\sigma(1)} \leq x_{\sigma(2)} \leq \dots \leq x_{\sigma(n)}$ ,  $X_{\sigma(i)} = \{\sigma(i), \sigma(i+1), \dots, \sigma(n)\}$  for  $i = 1, \dots, n$ . The resulting overall preference relation is therefore:

$$x \succsim y \text{ iff } S_v(x) \geq S_v(y).$$

This general aggregator has been introduced by Sugeno [Sugeno, 1974] in fuzzy sets theory and imported into decision theory under uncertainty where its use was axiomatically justified [Dubois et al., 1998]. The Sugeno integral can also be used in multicriteria decision making because the proposed axioms can easily be transposed. When  $v$  is a measure of possibility over  $N$  defined by  $v(A) = \max\{\pi_i, i \in A\}$ ,  $(\pi_1, \dots, \pi_n)$  playing the role of ordinal weights (positive coefficients such that  $\max\{\pi_i, i \in A\} = 1$ ), the Sugeno integral is nothing else but a *weighted maximum* defined by:

$$wmax(x) = \max_{i \in N} \min\{x_i, \pi_i\}$$

When  $v$  is a necessity measure defined on  $N$  by  $v(A) = 1 - \max\{\pi_i, i \notin A\}$ ,  $(\pi_1, \dots, \pi_n)$  playing the role of ordinal possibilistic weights, the Sugeno integral takes the particular form of a *weighted minimum* defined by:

$$wmin(x) = \min_{i \in N} \max\{x_i, 1 - \pi_i\}$$

The weighted max operator reflects an optimistic view which consists of valuing the existence of at least one good performance on an important criterion. The weighted min reflects a more pessimistic view which consists of assessing the extent to which no important criterion exists on which the alternative under consideration performs poorly. These two models as well as the Sugeno integral have been studied in depth in a decision-making framework, see e.g., [Dubois and Prade, 1995; Dubois et al., 2001b].

## 5 Conclusion

The models presented in this chapter provide an overview of the main aggregators used to take multiple points of view into account, whether in multi-criteria decision-making or collective decision-making (see also chapters I.15 and I.17). Most of them are widely used in artificial intelligence. In multicriteria and collective decision-making, the CA approach is very present through ordinal methods of aggregation derived from social choice theory and voting procedures. However, the AC approach remains the most widespread, whether in multi-criteria decision-making to determine specific trade-offs in the Pareto set (by optimizing a scalarizing function), or in collective decision-making to determine a fair Pareto-optimal solution. Concern-

ing this second approach, the reader wishing to obtain technical complements on the aggregation functions and their properties may refer to [Grabisch et al., 2009].

Whether they fall under the CA or AC approach, the models presented in this chapter can also be used in decision-making under uncertainty, when trying to evaluate and compare acts in the sense of Savage. If we consider situations in which uncertainty is represented by a finite set of possible states  $S = \{s_1, \dots, s_n\}$ , it appears that the different states act as different criteria to evaluate the possible acts. This explains why several models introduced in this chapter may also be used for decision under uncertainty. These criteria can even be generalized in the case of a continuous set of states of nature. The next chapter is precisely intended to present in the most general case the decision models used under uncertainty and risk.

The main theoretical axes that still need to be developed in multicriteria analysis relate to the axiomatic justification of existing models (the results to characterize the preferences that can be represented by a particular model do not always exist in multicriteria analysis, even if neighboring results sometimes exist for decision making under uncertainty), the development of decision models with increased descriptive power in the presence of rich information, and the development of ordinal or partial aggregation methods in the presence of poor information. From a more operational point of view, the main challenges of multicriteria decision theory are to propose efficient methods for eliciting or learning the parameters of the models they propose (for example for recommendation systems) and, on the other hand, to develop efficient algorithms for determining preferred solutions in combinatorial problems (preference-based search). The combinatorial nature of the space of feasible solutions precludes the use of any explicit enumeration method to compare solutions. The search for preferred solutions necessarily involves the development of implicit enumeration methods, but the optimization problems that need to be solved are all the more difficult as the models are sophisticated. Decision theory, by producing models that are always richer to account for various decision-making behaviors, is therefore a source of permanent challenges for computer scientists. These aspects of elicitation and computation are widely studied in artificial intelligence and are the subject of numerous recent contributions to algorithmic decision theory.

## References

- Arrow, K. (1951). *Social choice and individual values*. John Wiley, New York.
- Bazgan, C., Hugot, H., and Vanderpooten, D. (2009). Implementing an efficient FPTAS for the 0-1 multi-objective knapsack problem. *European J. of Operational Research*, 198(1):47–56.
- Beliakov, G., Calvo, T., and James, S. (2015). Aggregation functions for recommender systems. In *Recommender Systems Handbook*, pages 777–808.
- Benabbou, N., Perny, P., and Viappiani, P. (2017). Incremental elicitation of choquet capacities for multicriteria choice, ranking and sorting problems. *Artificial Intelligence Journal*, 246:152–180.

- Borda, J.-C. (1781). *Mémoire sur les élections au scrutin*. Comptes rendus de l'Académie des sciences. traduit par Alfred de Grazia comme Mathematical Derivation of a election system, Isis, vol 44, pp 42–51.
- Bouveret, S. and Lang, J. (2005). Efficiency and envy-freeness in fair division of indivisible goods. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI'05)*.
- Bouyssou, D. and Marchant, T. (2013). Multiattribute preference models with reference points. *European Journal of Operational Research*, 229(2):470–481.
- Charnes, A., Cooper, W., and Ferguson, R. (1955). Optimal estimation of executive compensation by linear programming. *Management Science*, 1:138–151.
- Chassein, A. and Goerigk, M. (2015). Alternative formulations for the ordered weighted averaging objective. *Information Processing Letters*, 115(6):604–608.
- Chateauneuf, A. and Tallon, J.-M. (1999). Diversification, convex preferences and non-empty core in the Choquet expected utility model. *Economic Theory*, 19(3):509–523.
- Choquet, G. (1953). Theory of capacities. *Annales de l'Institut Fourier*, 5:131–295.
- Couceiro, M., Dubois, D., Prade, H., and Waldhauser, T. (2012). Decision-making with Sugeno integrals: DMU vs. MCDM. In *Proc. European Conf. on Artificial Intelligence (ECAI'12)*, pages 288–293.
- Diakonikolas, I. and Yannakakis, M. (2008). Succinct approximate convex Pareto curves. In *Proc. ACM-SIAM Symp. on Discrete Algorithms (SODA'08)*, pages 74–83.
- Dubois, D., Fargier, H., and Perny, P. (2003). Qualitative decision theory with preference relations and comparative uncertainty: An axiomatic approach. *Artificial Intelligence*, 148(1):219–260.
- Dubois, D., Marichal, J., Prade, H., Roubens, M., and Sabbadin, R. (2001a). The use of the discrete Sugeno integral in decision-making: A survey. *Int. J. of Uncertainty, Fuzziness and Knowledge-Based Systems*, 9(5):539–561.
- Dubois, D. and Prade, H. (1995). Possibility theory as a basis of qualitative decision theory. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI'95)*, pages 1924–1930.
- Dubois, D. and Prade, H. (1997). k-order additive discrete fuzzy measures and their representation. *Fuzzy Sets and Systems*, 92:167–189.
- Dubois, D., Prade, H., and Sabbadin, R. (1998). Qualitative decision theory with Sugeno integrals. In *Proc. Conf. on Uncertainty in Artificial Intelligence (UAI'98)*, pages 121–128.
- Dubois, D., Prade, H., and Sabbadin, R. (2001b). Decision-theoretic foundations of qualitative possibility theory. *European Journal of Operational Research*, 128:459–478.
- Dubus, J., Gonzales, C., and Perny, P. (2009). Choquet optimization using GAI networks for multiagent/multicriteria decision-making. In *Proceedings of the international conference on Algorithmic Decision Theory*, pages 377–389.
- Elkind, E. and Ismaili, A. (2015). Owa-based extensions of the chamberlin-courant rule. In *Proceedings of the 4th International Conference on Algorithmic Decision Theory*, pages 486–502.

- Fallah Tehrani, A., Cheng, W., Dembczynski, K., and Hüllermeier, E. (2012). Learning monotone nonlinear models using the Choquet integral. *Machine Learning*, 89(1-2):183–211.
- Fodor, J. and Roubens, M. (1994). *Fuzzy Preference Modelling and Multicriteria Decision support*. Kluwer Academic Publishers.
- Galand, L. and Perny, P. (2006). Search for compromise solutions in multiobjective state space graphs. In *Proceeding of the 17th European Conference on Artificial Intelligence*, pages 93–97.
- Galand, L. and Perny, P. (2007). Search for choquet-optimal paths under uncertainty. In *UAI 2007, Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence, Vancouver, BC, Canada, July 19-22, 2007*, pages 125–132.
- García-Lapresta, J. L. and Martínez-Panero, M. (2017). Positional voting rules generated by aggregation functions and the role of duplication. *International Journal on Intelligent Systems*, 32(9):926–946.
- Geoffrion, A., Dyer, J., and Feinberg, A. (1973). An interactive approach for multicriteria optimization with an application to the operation of an academic department. *Management Science*, 19:357–369.
- Golden, B. and Perny, P. (2010). Infinite order Lorenz dominance for fair multiagent optimization. In *Proc. Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS'10)*, pages 383–390.
- Goldsmith, J., Lang, J., Mattei, N., and Perny, P. (2014). Voting with rank dependent scoring rules. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 698–704.
- Grabisch, M. (1996). The application of fuzzy integrals in multicriteria decision making. *European J. of Operational Research*, 89(3):445–456.
- Grabisch, M. (2016). *Set Functions, Games and Capacities in Decision Making*. Springer.
- Grabisch, M. and Labreuche, C. (2008). A decade of application of the Choquet and Sugeno integrals in multi-criteria decision aid. *4OR*, 6(1):1–44.
- Grabisch, M., Marichal, J., and Mesiar, R. (2009). *Aggregation Functions*. Cambridge University Press.
- Grabisch, M., Nguyen, H., and Walker, E. (1995). *Fundamentals of Uncertainty Calculi, with Applications*. Encyclopedia of Mathematics and its Applications. Kluwer Academic Publishers.
- Gábor, Z., Kalmár, Z., and Szepesvári, C. (1998). Multicriteria reinforcement learning. In *Proc. Int. Conf. of Machine Learning (ICML'98)*.
- Hansen, P. (1980). Bicriterion path problems. In Fandel, G. and Gal, T., editors, *Int. Conf. on Multiple Criteria Decision Making (MCDM'80)*.
- Hüllermeier, E. and Fallah Tehrani, A. (2013). Efficient learning of classifiers based on the 2-additive Choquet integral. In *Computational Intelligence in Intelligent Data Analysis Studies in Computational Intelligence Volume*, volume 445, pages 17–29.

- Laumanns, M., Thiele, L., Deb, K., and Zitzler, E. (2002). Combining convergence and diversity in evolutionary multiobjective optimization. *Evolutionary Computation*, 10(3):263–282.
- Lesca, J., Minoux, M., and Perny, P. (2013). Compact versus noncompact LP formulations for minimizing convex choquet integrals. *Discrete Applied Mathematics*, 161(1-2):184–199.
- Lesca, J. and Perny, P. (2010). LP sovable models for multiagent fair allocation problems. In *Proc. European Conf. on Artificial Intelligence (ECAI'10)*, pages 387–392.
- Lovász, L. (1983). Submodular functions and convexity. In Bachem, A., Grötschel, M., and Korte, B., editors, *Mathematical Programming, the State of the Art*, pages 235–257.
- Marichal, J.-L. (2000a). An axiomatic approach of the discrete Choquet integral as a tool to aggregate interacting criteria. *IEEE Transactions on Fuzzy Systems*, 8(6):800–807.
- Marichal, J.-L. (2000b). On Sugeno integral as an aggregation function. *Fuzzy Sets and Systems*, 114(3).
- Marichal, J.-L. and Roubens, M. (2000). Determination of weights of interacting criteria from a reference set. *European J. of Operational Research*, 124(3):641–650.
- Marquis de Condorcet (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Imprimerie Royale, Paris.
- Marshall, W. and Olkin, I. (1979). *Inequalities: Theory of Majorization and its Applications*. Academic Press, London.
- Moulin, H. (1988). *Axioms of cooperative decision making*. Monograph of the Econometric Society. Cambridge University Press.
- Muliere, P. and Scarsini, M. (1989). A note on stochastic dominance and inequality measures. *J. of Economic Theory*, 49:314–323.
- Ogryczak, W. and Sliwinski, T. (2003). On solving linear programs with the ordered weighted averaging objective. *European J. of Operational Research*, 148(1):80–91.
- Ogryczak, W. and Sliwinski, T. (2007). On optimization of the importance weighted OWA aggregation of multiple criteria. In *Proc. Int. Conf. on Computational Science and Its Applications (ICCSA'07)*, volume 4705 of *Lecture Notes in Computer Science*, pages 804–817.
- Papadimitriou, C. and Yannakakis, M. (2000). On the approximability of trade-offs and optimal access of web sources. In *Proc. IEEE Symp. on Foundations of Computer Science (FOCS'00)*, pages 86–92.
- Pareto, V. (1906). *Manuale di Economia Politica*. Piccola Biblioteca Scientifica, Milan. Traduit en anglais par Ann S. Schwier (1971), *Manual of Political Economy*, MacMillan, London.
- Perny, P. (1998). Multicriteria filtering methods based on concordance and non-discordance principles. *Annals of Operations Research*, 80:137–165.

- Perny, P. and Rolland, A. (2006). Reference-dependent qualitative models for decision making under uncertainty. In *Proc. European Conf. on Artificial Intelligence (ECAI'06)*, pages 422–426.
- Perny, P. and Roy, B. (1992). The use of fuzzy outranking relations in preference modelling. *Fuzzy Sets and Systems*, 49:33–53.
- Perny, P. and Spanjaard, O. (2008). Near admissible algorithms for multiobjective search. In *Proc. European Conf. on Artificial Intelligence (ECAI'08)*, pages 490–494.
- Perny, P., Viappiani, P., and Boukhatem, A. (2016). Incremental preference elicitation for decision making under risk with the rank-dependent utility model. In *Proceedings of UAI'16*, pages 597–606.
- Pirlot, M. and Vincke, P. (1997). *Semiorders Properties, Representations, Applications*. Kluwer Academic Publishers.
- Rolland, A. (2013). Reference-based preferences aggregation procedures in multi-criteria decision making. *European Journal of Operational Research*, 225(3):479–486.
- Roubens, M. and Vincke, P. (1985). *Preference Modelling*. Springer-Verlag, Berling.
- Roy, B. (1985). *Méthodologie multicritère d'aide à la décision*. Economica.
- Roy, B. and Bouyssou, D. (1993). *Méthodologie multicritère d'aide à la décision : méthodes et cas*. Economica.
- Schiex, T., Fargier, H., and Verfaillie, G. (1995). Valued constraint satisfaction problems: Hard and easy problems. In *Proc. Int. Joint Conf. on Artificial Intelligence (IJCAI'95)*, pages 631–639.
- Sen, A. (1986a). Social choice theory. In Arrow, K. and Intriligator, M., editors, *Handbook of mathematical economics*, volume 3, pages 1073–1181. North-Holland, Amsterdam.
- Sen, A. K. (1986b). Social choice theory. In Arrow, K. and Intriligator, M., editors, *Handbook of Mathematical Economics*, chapter 22, pages 1073–1181. Elsevier Sciences Publishers, North-Holland.
- Shorrocks, A. (1983). Ranking income distributions. *Economica*, 50:3–17.
- Skowron, P., Faliszewski, P., and Lang, J. (2016). Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artificial Intelligence Journal*, 241:191–216.
- Steuer, R. and Choo, E.-U. (1983). An interactive weighted Tchebycheff procedure for multiple objective programming. *Mathematical Programming*, 26:326–344.
- Steuer, R. E. (1986). *Multiple Criteria Optimization: Theory, Computation and Application*. John Wiley, New York.
- Stewart, B. S. and White III, C. C. (1991). Multiobjective A\*. *J. of the Association for Computing Machinery*, 38(4):775–814.
- Sugeno, M. (1974). *Theory of fuzzy integrals and its applications*. PhD thesis, Tokyo Institute of Technology.
- Torra, V. (1997). The weighted OWA operator. *Int. J. of Intelligent Systems*, 12:153–166.

- Torra, V. and Narukawa, Y. (2007). *Modeling decisions - information fusion and aggregation operators*. Springer.
- Vincke, P. (1992). *Multicriteria Decision Aid*. Wiley.
- Weymark, J. (1981). Generalized Gini inequality indices. *Mathematical Social Sciences*, 1:409–430.
- White, D. (1982). Multi-objective infinite-horizon discounted Markov decision processes. *J. of Mathematical Analysis and Applications*, 89:639–647.
- Wierzbicki, A. (1986). On the completeness and constructiveness of parametric characterizations to vector optimization problems. *OR Spektrum*, 8:73–87.
- Wierzbicki, A. P. (1999). Reference point approaches. In Gal, T., Stewart, T., and Hanne, T., editors, *Multicriteria Decision Making: Advances in MCDM models, Algorithms, and Applications*, chapter 9. Kluwer Academic Publishers.
- Yaari, M. (1987). The dual theory of choice under risk. *Econometrica*, 55:95–115.
- Yager, R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. In *IEEE Trans. on Systems, Man and Cybernetics*, volume 18, pages 183–190.



# Index

- aggregation, 4
- capacity, 21
- Choquet integral, 21
- concordance, 13
- Condorcet effect, 8
- covering
  - $\varepsilon$ -covering, 9
- criterion, 2
- decision
  - multicriteria, 2
- dictatorial procedure, 9
- dominance
  - $\varepsilon$ -dominance, 9
  - Lorenz, 10
  - weighted, 12
  - partial, 9
  - weak Pareto, 9
- Möbius inverse, 23
- model
  - Yaari, 19
- multicriteria, 2
- norm
  - Tchebycheff, 16
- ordered average
  - OWA, 18
  - WOWA, 19
- Pareto dominance, 3
- performance, 2
- Pigou-Dalton transfers, 10
- point
  - ideal, 17
  - reference, 14
- preference aggregation, 4
- problem
  - choice, 3
  - ranking, 3
  - sorting, 3
- relation
  - concordance, 13
- second order stochastic dominance, 12
- Sugeno integral, 24
- veto, 14
- Weak Pareto dominance, 3