European Journal of Operational Research 311 (2023) 679–689

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Analytics, Computational Intelligence and Information Management

The transport problem for non-additive measures

Vicenç Torra

Department of Computing Sciences, Umeå University, Sweden

ARTICLE INFO

Article history:
Received 7 December 2022
Accepted 10 March 2023
Available online 16 March 2023

Keywords:
Fuzzy sets
Non-additive measures
Fuzzy measures
Optimal transport

ABSTRACT

Non-additive measures, also known as fuzzy measures, capacities, and monotonic games, are increasingly used in different fields. Applications have been built within computer science and artificial intelligence related to e.g., decision making, image processing, machine learning for both classification, and regression. Tools for measure identification have been built. In short, as non-additive measures are more general than additive ones (i.e., than probabilities), they have better modeling capabilities allowing to model situations and problems that cannot be modeled by the latter. See e.g., the application of non-additive measures and the Choquet integral to model both Ellsberg paradox and Allais paradox. Because of that, there is an increasing need to analyze non-additive measures.

The optimal transport problem is a useful tool for probabilities. It is used to define the Wasserstein distance, and also to solve problems as data matching. In this work with tackle the problem of defining the optimal transport problem for non-additive measures. We consider that it is necessary to provide appropriate definitions with a flavour similar to the one for probabilities, but for non-additive measures. The solutions should generalize the classical ones.

We provide definitions based on the Möbius transform, but also based on the $(\max, +)$-transform that we consider that has some advantages. We will discuss in this paper the problems that arise to define the transport problem for non-additive measures, and provide ways to solve them. We include the definitions of the optimal transport problem, and prove some properties. We also conclude providing a definition of the Wasserstein distance for non-additive measures.

© 2023 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

1. Introduction

The optimal transport (Bogachev & Kolesnikov, 2012; Santambrogio, 2015; Villani, 2003; 2008) is a well known problem that has been studied from a theoretical perspective and it is currently extensively used in applications. The optimal transport problem was introduced by Kantorovich (1942) and is a generalization of Monge's optimal transport problem. Kantorovich's optimal problem can be seen as establishing a relationship – an assignment – between two probability spaces. This relationship is optimal with respect to an underlying cost function. The cost function combined with the optimal assignment can then be used to define a distance between the two probability distributions. It is the Wasserstein distance. There are a large number of applications (Santambrogio, 2015) of the optimal transport and the Wasserstein distance in statistics and machine learning. For example, the Wasserstein GANs (Arjovsky et al., 2017).

Non-additive measures (Torra et al., 2013a), also known as fuzzy measures (Sugeno, 1972; 1974) and monotonic games, generalize probabilities by replacing the additivity condition by a less restrictive one. More particularly, they just require measures to be monotonic with respect to set inclusion. Distorted probabilities (Chateauneuf, 1996; Edwards, 1953; Honda et al., 2002a; Honda et al. 2002b; Honda & Narukawa, 2014) are an example of these measures. Non-additive measures are used in economics, decision making, and artificial intelligence. They permit to model situations that cannot be modeled with probability measures. For example, both Ellsberg paradox and Allais paradox, which correspond to decision problems, cannot be represented with probabilities but they are representable (solved) using non-additive measures (Gilboa, 2009). Similarly, some classification and regression problems (Tehrani et al., 2012) can be better solved with non-additive measures because they have additional degrees of freedom than additive ones (probabilities).

A fundamental difference between additive and non-additive measures is that the former consider the elements of the singletons independent while this is not so for non-additive ones. That

E-mail address: vtorra@cs.umu.se

https://doi.org/10.1016/j.ejor.2023.03.016
0377-2217/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

is, for a non-additive measure μ on $X = \{x_1, x_2, \ldots, x_n\}$ we may have $\mu(\{x_1, x_2\}) > \mu(\{x_1\}) + \mu(\{x_2\})$ if we have a positive interaction of x_1 and x_2, or we may have $\mu(\{x_1, x_2\}) < \mu(\{x_1\}) + \mu(\{x_2\})$ if we have a negative interaction of x_1 and x_2. For probabilities, only equality is possible. See e.g., Gilboa (2009); Torra et al. (2018) for more concrete examples on what can be modeled with non-additive measures and cannot be modeled with probabilities. Then, in applications (Bardozzo et al., 2021; Beliakov, 2022; Marco-Detchart et al., 2021; Pereira et al., 2020), the measures represent some background knowledge on the variables or attributes, and the non-additive integrals (Benvenuti et al., 2002; Mesiar & Mesiarová, 2004; Narukawa et al., 2016; Torra et al., 2013a) are used to aggregate or combine the data with respect to the measure. The two most used non-additive integrals are the Choquet (1953) and Sugeno (1974) integrals, but there are several rich families of integrals (see e.g., Dimuro et al., 2020a; Dimuro et al., 2020b; Mesiar et al., 2016).

The more non-additive measures are used in applications, the more we need tools to compare and assess them. This naturally includes a need for similarities, distances, metrics, and divergences between pairs of measures. That is, we need a way to know when two measures are similar or not, and the extent of this similarity. As a more concrete example. In Choquistic regression (Tehrani et al., 2012) and in Choquistic-based metric learning (Abril et al., 2012) we determine non-additive measures from data. Given slightly different datasets the learned measure will differ. Which is the extent of this difference? How different are these measures for different problems? Similarly, we can use non-additive measures to model decisions (Gilboa, 2009; Mayag & Bouyssou, 2020). Then, different ways to make decisions are coded in the measures. Again, how to evaluate the difference between these measures.

Previous research on distances for non-additive measures exist. In particular, there are approaches to extend f-divergences through Radon-Nikodym-like derivatives. We have contributed (Torra et al., 2013b; 2014; 2016; 2020) to some of these results. See also the works by Agahi (2020); Agahi & Yadollahzadeh (2021) and by Ontkovičová & Kiselák (2022). The definition of f-divergences permits us to study Kullback–Leibler divergence, and define entropy, as well as the Principle of Minimum Discrimination and of Maximum Entropy (Torra, 2017). This links with previous research on entropy for measures (Honda & Narukawa, 2014). The Wasserstein distance is a well established distance for probabilities and, in general, for additive measures. It is based on the optimal transport problem. So, for its definition we need first to introduce the optimal transport problem for non-additive measures.

In fact, the optimal transport problem for additive measures has interesting mathematical properties and has been extensively used in applications. Properties (Ambrosio et al., 2008; Santambrogio, 2015) includes dual optimization problems, and the definition of gradient flows. For some families of cost functions, properties can be inferred on the result. E.g., that the Wasserstein distance is a proper distance satisfying the triangle inequality (e.g., Euclidean distance as cost function). In addition of its use to define the Wasserstein distance, the optimal transport problem has been used in other applications as data alignment (Huizing et al., 2022). Therefore, it seems a relevant problem to study it for non-additive measures as well. We study and formalize this problem in this paper. We can also solve it in practice for measures on a finite set defining an appropriate optimization problem, which is a linear optimization problem with linear constraints. Up to our knowledge no solution has been provided in the literature. This is so because of the difficulty to deal with the measure on non-singletons and to deal with positive and negative interactions on these sets. Let us briefly discuss some of these difficulties. Additional details and appropriate formalization will be given later in the paper.

- Transport problems can be understood as the transfer of mass (i.e., probability) from one object to another one. In probabilities, transfer is between elements of a reference set. Then, for non-additive measures, we need to take into account, not only the transfer from singletons to singletons, but also from and to sets of arbitrary cardinality.
- The mass associated to a non-singleton A is not independent but, due to the monotonicity condition, naturally depends on the mass associated to singletons for $x \in A$ and, in general, to any subset A' of the set A. Any definition of mass transfer needs to take this characteristic into account. In other words, it may be inappropriate to consider, in a solution, assigning mass associated to $\{x_1\}$ and $\{x_2\}$ to a set A and assigning the mass of $\{x_1, x_2\}$ to a set B such that $A \cap B = \emptyset$.
- In a way, the total mass associated to two non-additive measures μ and ν is not necessarily the same even when $\mu(X) = \nu(X) = 1$. For example, while in a probability, the total probability assigned to singletons is one, this is not necessarily true for non-additive measures. Note that it can be any number in the interval $[0, n]$, where n is the cardinality of X. This adds an additional complexity into the process. This problem has some connections with the case of unbalanced transport problems studied in the optimal problem transport community. They are problems in which two measures do not have the same total mass.

It is relevant to underline that for non-additive measures, the family of distances and similarities related to the f-divergence are based on the Choquet integral (i.e., an integral with respect to a non-additive measure). Their Radon-Nikodym-like derivative briefly mentioned above is defined taking into account such integral. In contrast, the definitions of the transport problem discussed in this paper, do not depend on this integral and is based on standard integration (i.e., addition in the discrete case).

- The transport problem considers costs and assignments, and the addition of their product to evaluate a solution. The integral is used when the domains are continuous. If we consider non-additive measures, an approach can consist on using non-additive integrals when computing this cost. Nevertheless, the Choquet integral does not satisfy, in general, Fubini's theorem. This approach is sketched in Gal & Niculescu (2019). Nevertheless, it is difficult to solve this approach in practice, and the projection of the measures on the product space into each of the original spaces would need to be clearly defined. The results we provide in this work can help on this definition.

The structure of the paper is as follows. In Section 2 we review the concepts we will use in the rest of the paper. In Section 3 we introduce our definitions of the transport problem for non-additive measures. We present our main results and discuss the relationship with the standard transport problem. The paper finishes with some research directions for future work.

2. Preliminaries

We divide this section in two parts. One reviewing the transport problem and the other reviewing some definitions we need related to non-additive measures. In what follows, we consider measures on finite reference sets.

2.1. The transport problem

Let us consider two additive measures P and Q on X and Y, respectively, with probability distributions p and q. Then, the transport problem (Santambrogio, 2015; Villani, 2003; 2008) consists on

finding an assignment γ from p to q. The assignment γ needs to have as marginals p and q. That is, $p(x) = \sum_{y\in Y} \gamma(x,y)$ and $q(y) = \sum_{x\in X} \gamma(x,y)$. An assignment γ that satisfies the marginals and is positive is called a feasible assignment.

Then, given a cost function $c : X \times Y \to \mathbb{R}^+$, the optimal transport problem corresponds to the assignment that minimizes the total cost, where this total cost is defined by

$$\sum_{x\in X}\sum_{y\in Y} c(x,y)\gamma(x,y)$$

over the space of all possible assignments γ with marginals p and q.

It is usual to consider the transport problem for pairs of measures on the same reference set. That is, $X = Y$. In this case, the cost function is $c : X \times X \to \mathbb{R}^+$ and typically $c(x,x) = 0$ for all $x \in X$. This is a requirement when we define Wasserstein distance (or Kantorovich–Rubinstein metric) in terms of the optimal transport problem. More formally, the Wasserstein distance is defined as

$$d(P,Q) = \inf_{\gamma\in\Gamma(P,Q)} \sum_{x\in X}\sum_{y\in Y} c(x,y)\gamma(x,y), \tag{1}$$

where $c(x,y) = |x-y|$.

Note that other cost functions are also used.

2.2. Non-additive measures

We begin this review defining a measure on a reference set X. We consider that this set is finite. Details on non-additive measures can be found in several reference works (Chateauneuf & Jaffray, 1989; Denneberg, 1994; Torra et al., 2013a).

Definition 1. Given a finite reference set X, a set function $\mu : 2^X \to [0,1]$ is a non-additive measure if:

- $\mu(\emptyset) = 0$ (boundary condition)
- If $A \subseteq B$ then $\mu(A) \leq \mu(B)$ for $A, B \subseteq X$.

Observe that non-additive measures are also known as capacities, fuzzy measures, and monotonic games.

If the measure is such that $\mu(X) = 1$ we say that the measure is normalized.

There are alternative ways to represent non-additive measures. The Möbius transform is one of them. We review its definition. We will also present the $(\max, +)$-transform. The latter transform has some connections with the generalizations (Mesiar, 1997; Mesiar, 1999) of k-order additive measures and Möbius transform. See Torra (2022) for details.

Definition 2. Let μ be a non-additive measure on X. Then, its Möbius transform τ_μ corresponds to:

$$\tau_\mu(A) = \sum_{B\subseteq A} (-1)^{|A|-|B|}\mu(B).$$

In this definition we use $\subseteq$ to denote the non-strict inclusion, and $\subset$ to denote the strict inclusion.

Then, given a function τ_μ that is the Möbius transform of μ, we have that for all $A \subseteq X$:

$$\mu(A) = \sum_{B\subseteq A} \tau_\mu(B).$$

In general, a set function m on X is a Möbius transform of a monotone measure if

- (i) $m(\emptyset) = 0$,
- (ii) $\sum_{B\subseteq A} m(B) \geq 0$ for all $A \subseteq X$, and
- (iii) for $A \subseteq A'$ it holds $\sum_{B\subseteq A} m(B) \leq \sum_{B\subseteq A'} m(B)$.

We can also add a condition for the normalization of the resulting measure. This naturally corresponds to require $\sum_{A\subseteq X} m(A) = 1$.

A well known class of measures are the so-called belief functions. They are characterized by the fact that the Möbius transform is always non-negative and adds to one (i.e., $m(A) \geq 0$) and $\sum_{A\subseteq X} m(A) = 1$. In this case, the Möbius transform is called a basic probability assignment (bpa). It is easy to see that for singletons $\{x_i\}$, the measure corresponds to the basic probability assignment or Möbius transform of μ. That is, i.e., $\mu(\{x_i\}) = m(\{x_i\})$.

We discuss now an alternative transform. As we have stated above, there are some connections between this transform and works by Mesiar generalizing the Möbius transform (see the works introducing the $\oplus$-Möbius transform Mesiar, 1997; Mesiar, 1999).

Definition 3 (Torra (2022))**.** Let μ be a non-additive measure on X. Then, we define the $(\text{Max}, +)$-transform as the set function $\tau_\mu : 2^X \to \mathbb{R}^+$ such that:

$$\tau_\mu(B) = \mu(B) - \max_{A\subset B} \mu(A) \tag{2}$$

It can be seen from this definition that τ_μ is always positive and that for an arbitrary normalized measure μ and any set $A \subseteq X$, the $(\max, +)$-transform $\tau_\mu(A)$ is at most one. That is, $\tau_\mu(A) \in [0,1]$ for all $A \subseteq X$.

When μ is an additive measure, the following proposition holds. We will use this result later to study the transport problem when the measures are additive.

Proposition 4. *Let X be a reference set, and μ be a measure that is additive. Then, if τ_μ is the $(\max, +)$-transform of μ, the following holds:*

$$\tau_\mu(B) = \min_{x_i\in B} \mu(\{x_i\}).$$

Proof. If μ is additive, then $\max_{A\subset B} \mu(A)$ will be achieved with the set A_0 containing $|A_0| = |A| - 1$ elements x_i. These elements $x_i \in A$ are the ones with the largest values $\mu(\{x_i\})$. Therefore, the one that is missing in A_0 is the one x_0 with the smallest value $\mu(\{x_0\})$. Thus, $\mu(B) - \mu(A_0) = \mu(\{x_0\}) = \min_{x_i\in B} \mu(\{x_i\})$. □

We can also prove the following.

Proposition 5 (Torra (2022))**.** *Let m be a set function over the set X such that $m : 2^X \to [0,1]$, with $m(\emptyset) = 0$. Then, the set function defined by*

$$\mu(B) = \left(\max_{A\subset B} \mu(A)\right) + m(B), \tag{3}$$

is a non-additive (normalized) measure and its $(\text{Max}, +)$-transform is m.

Observe that, in general, for an arbitrary positive set function m as above, the resulting measure is not necessarily normalized. For a given μ, and its corresponding $(\text{Max}, +)$-transform m, we have that Proposition 5 returns μ as the measure associated to m.

3. Definitions

We will provide now alternative definitions for the transport problem. We begin using the basic probability assignment, which provides the definition that is closest to standard transport problems for probability distributions. We will show the limitations of this approach, and the difficulties to extend it to other types of non-additive measures. It will also provide a direction for other approaches.

We will consider, in general, two non-additive measures μ and ν, both defined on the same reference set X. The definitions we provide for the transport problem are easily generalized to the case of μ and ν defined on different reference sets X and Y.

3.1. Beliefs functions and basic probability assignments

We begin considering the transport problem for belief measures. That is, we will consider two belief measures μ and ν on the reference set X. As we have seen above, for belief measures, the Möbius transform is called a basic probability assignment (bpa) and is a function m such that $m(A) \geq 0$ for all A. As, in addition, $\sum_{A\subseteq X} m(A) = 1$, basic probability assignments can be understood as probability distributions on the power set of X. Therefore, we can apply the definition of the optimal transport problem using as the reference set the power set of our original reference set. That is, 2^X. This type of definition appears in Bronevich & Rozenberg (2021).

Definition 6. Let μ and ν be two belief functions with τ_μ and τ_ν be the corresponding Möbius transforms. That is, τ_μ and τ_ν are basic probability assignments and, thus, they are positive, add to one, and are zero in the empty set. Then, given a cost function $c_b : 2^X \times 2^X \to \mathbb{R}^+$ we define the corresponding transport problem as follows.

Find the assignment $assg : 2^X \times 2^X \to [0, 1]$ that minimizes the following objective function:

$$OF = \sum_{A\subseteq X}\sum_{B\subseteq X} c_b(A, B)assg(A, B) \tag{4}$$

In the definition, $assg$ is a function $assg : 2^X \times 2^X \to [0, 1]$, and the marginals of this function need to correspond to τ_μ and τ_ν. In other words, $\tau_\mu(A) = \sum_{B\subseteq X} assg(A, B)$ and $\tau_\nu(B) = \sum_{A\subseteq X} assg(A, B)$.

We want to stress that in this definition, A or B can be the empty set, but this does not play any role as the basic probability assignments associated to the empty sets are zero for both μ and ν, and the assignment is non-negative. That is, $\tau_\mu(\emptyset) = \tau_\nu(\emptyset) = 0$ and $assg(\emptyset, B) = assg(A, \emptyset) = 0$ for all $A \subseteq X$ and $B \subseteq X$.

Let us consider some properties. The proof of the first one is trivial from the definition above.

Proposition 7. *Let μ and ν as above, and let assg be as in the previous definition; then, $\sum_{A\subseteq X}\sum_{B\subseteq X} assg(A, B) = 1$.*

Proposition 8. *The following two conditions are equivalent*

- $\sum_{B\subseteq X} assg(A, B) = \tau_\mu(A)$ *for all* $A \subseteq X$, *and*
- $\mu'(A) = \sum_{A'\subseteq A}\sum_{B\subseteq X} assg(A', B) = \mu(A)$ *for all* $A \subseteq X$.

These conditions are given focusing on $\tau_\mu(A)$. A similar proposition can be proven when we consider $\tau_\nu(B)$ for all $B \subseteq X$.

Proof. To prove this proposition, we prove

- If $\sum_{B\subseteq X} assg(A, B) = \tau_\mu(A)$ for all $A \subseteq X$, then $\mu'(A) = \sum_{A'\subseteq A}\sum_{B\subseteq X} assg(A', B) = \mu(A)$ for all $A \subseteq X$,
- If $\mu(A) = \sum_{A'\subseteq A}\sum_{B\subseteq X} assg(A', B)$ for all $A \subseteq X$, then $\sum_{B\subseteq X} assg(A, B) = \tau_\mu(A)$ for all $A \subseteq X$.

The first implication is trivial. Let us define $\mu'(A) = \sum_{A'\subseteq A} \tau_\mu(A')$ which is naturally equivalent to $\mu(A)$ and then replace $\tau_\mu(A')$ by its corresponding expression on the left.

The second implication can be proven by induction. First we consider it for a set of cardinality 1. That is, we consider $A = \{x_i\}$. In this case the only A' to consider in the summatory is also $A' = \{x_i\}$. So, the expression in the proof corresponds to $\mu(\{x_i\}) = \sum_{B\subseteq X} assg(\{x_i\}, B)$ which is naturally equivalent to $\tau_\mu(\{x_i\})$. Then, we consider in the induction hypothesis that the condition is true for sets of cardinality n smaller than the cardinality of X (i.e., $n < |X|$). Let us now prove that it is also true for a set of cardinality $n + 1$. That is, we consider $\mu(A) = \sum_{A'\subseteq A}\sum_{B\subseteq X} assg(A', B)$ and rewrite it distinguishing A from the other sets different to A. That is, $\mu(A) = \sum_{A'\subset A}\sum_{B\subseteq X} assg(A', B) + \sum_{B\subseteq X} assg(A, B)$. By the induction hypothesis, we have that $\sum_{B\subseteq X} assg(A', B) = \tau_\mu(A')$ therefore, as $\mu(A) = \sum_{A'\subseteq A} \tau_\mu(A')$,

$$\mu(A) = \sum_{A'\subseteq A} \tau_\mu(A') = \sum_{A'\subset A} \tau_\mu(A') + \sum_{B\subseteq X} assg(A, B).$$

So, $\sum_{B\subseteq X} assg(A, B)$ should be $\tau_\mu(A)$, and the proposition is proven. □

This proposition shows that both conditions can be equivalently used in the transport problem to denote the marginals of the measure. That is, to restrict the values of the marginals of the assignment we can use either the basic probability assignment or the measure itself. We prefer to use the constraint using the basic probability assignment, as we have written in Definition 6. In this way we need to consider less terms in the constraint, and it is simpler to interpret. The constraint for each $A \subset X$ corresponds to a row in the matrix that represents the assignment for pairs (A, B).

3.2. Möbius transform

When the measure is not a belief function, the Möbius transform will contain some negative values. It is important to note that we not only need to deal with negative values, but the values can be arbitrarily large, for an appropriate number of elements in the reference set. Let us illustrate this with an example.

Example 9. Let μ be a non-additive measure defined as zero for all sets of cardinality smaller than n, and 1 for all sets of cardinality at least n. That is, $\mu(A) = 1$ if and only if $|A| \geq n$ (and $\mu(A) = 0$ otherwise).

Then, its Möbius transform will be zero for all sets of cardinality smaller than n, and will be one for all sets of cardinality n. In addition, for any set of cardinality $n + 1$ there will be $n + 1$ subsets with cardinality n and, so, its Möbius transform will be $\tau_\mu(A) = \mu(A) - (n + 1) = 1 - (n + 1) = -n$.

The transform will be still larger for the sets of cardinality $n + 2$. In this case there are $n + 2$ sets of cardinality $n + 1$ and $(n + 2)(n + 1)/2$ sets of cardinality n. Therefore,

$$\tau_\mu(A) = 1 - (n + 2)(-n) + \frac{(n + 2)(n + 1)}{2} = \frac{n^2 + n}{2}.$$

The transport problem as given in Definition 6 is, in general, unsuitable for assignments that can be negative (i.e., $assg : 2^X \times 2^X \to \mathbb{R}$). It is easy to see that for any cost function such that $c_b(A, A) = c_b(B, B) = 0$ and $c_b(A, B) = c_b(B, A) = \kappa > 0$ and a feasible assignment $assg$, we can define a new feasible assignment $assg' = assg + assg_\alpha$ with $assg_\alpha(A, A) = assg_\alpha(B, B) = +\alpha$ and $assg_\alpha(A, B) = assg_\alpha(B, A) = -\alpha$. Note that this assignment is feasible because $assg(A, A) + assg(A, B) = 0 + \kappa = assg'(A, A) + assg'(A, B) = 0 + \kappa + \alpha - \alpha$. Similarly, $assg(B, A) + assg(B, B) = 0 + \kappa = assg'(B, A) + assg'(B, B) = 0 + \kappa - \alpha + \alpha$. This assignment will increase arbitrarily the objective function. Taking these considerations into account it is natural to consider the absolute value of the assignment. This results into the following objective function. This objective function replaces the one in Definition 6 (Eq. (4)). We use here c_M to denote the cost function (to distinguish it from the one for basic probability assignments c_b.

$$OF = \sum_{\emptyset\subset A\subseteq X}\sum_{\emptyset\subset B\subseteq X} c_M(A, B)|assg(A, B)| \tag{5}$$

The assignment $assg$ needs to satisfy the marginals, either positive or negative, about the Möbius transform. Observe that here the equivalences on the expressions in Proposition 8 also apply.

We have shown in Example 9 that we can have a Möbius transform with very large values. We will now consider two measures similar to Example 9 and show its effect in the value of the objective function. The example will illustrate that, in a way, the mass is counted multiple times in the cost function.

Example 10. Let μ and ν be two non-additive measures defined on the reference set $X = \{x_1, \ldots, x_n, x_{n+1}, \ldots, x_{3n}\}$. Let $A_0 = \{x_1, \ldots, x_n, x_{n+1}\}$, and let $B_0 = \{x_{n+2}, \ldots, x_{2n+2}\}$.

Let us define μ as follows. Let $\mu(A) = 1$ for all subsets of A_0 with cardinality n. Let $\mu(A) = 0$ for all subsets of A_0 with cardinality smaller than n. Let $A_i = A_0 \setminus \{x_i\}$ for $i = 1, \ldots, n+1$. Therefore, for these sets $\mu(A_i) = 1$, as they have cardinality n. In addition, all supersets of these sets will also have measure one.

Similarly, let $\nu(B) = 1$ for all subsets of B_0 with cardinality n, and $\nu(B) = 0$ for all subsets of B_0 with cardinality smaller than n. Let $B_i = B_0 \setminus \{x_{n+1+i}\}$ for $i = 1, \ldots, n+1$, so $\nu(B_i) = 1$. Similarly, all supersets of these sets will also have measure one.

Let us now consider a cost function. For simplicity, the cost function is $c_M(A, B) = 1$ for $A \neq B$ and $c_M(A, B) = 0$ for $A = B$.

Then, a natural assignment is to assign $assg(A_i, B_i) = 1$ for $i = 1, \ldots, n+1$. Nevertheless, we also need to assign the mass for the other sets with non zero Möbius transform. Note that the Möbius transform for both A_0 and B_0 will be $-n$. So, we need e.g., $assg(A_0, B_0) = -n$. We will have larger sets with larger assignments. E.g., $assg(A_0 \cup \{x_{3n}\}, B_0 \cup \{x_{3n-1}\}) = (n^2 + n)/2$.

This example illustrates that the definition of the problem in terms of the Möbius transform causes the mass of some sets appear more than once in the cost. I.e., the value of $-n$ associated to A_0 mainly corresponds to the mass of all A_i which we also need to assign. We have assigned it to B_i with $assg(A_i, B_i)$. In other words, we have transferred one unit from A_1 to B_1 with $assg(A_1, B_1) = 1$, and the same from A_i to B_i, and then we need to transfer this mass again through the assignment $assg(A_0, B_0) = -n$.

The formulation of the optimal transport using the $(\max, +)$-transform mitigates this problem. This is discussed in the next section. As we will see, this is at the cost of having unassigned mass.

3.3. $(\max, +)$-transform

The two definitions above focus on the Möbius transform. We now consider the $(\max, +)$-transform.

A first thought is to use the same approach as in Definition 6 but using now the $(\max, +)$-transform. That is, use Eq. (4) and the marginals that correspond to $\tau_\mu(A)$ and $\tau_\nu(B)$ where these expressions correspond to:

$$\tau_\mu(A) = \sum_{\emptyset \subset B' \subseteq X} assg(A, B')$$

$$\tau_\nu(B) = \sum_{\emptyset \subset A' \subseteq X} assg(A', B)$$

Unfortunately, this approach does not always lead to a solution. That is, there are pairs of measures μ and ν for which there is no feasible assignment. This is illustrated in the following example.

Example 11. Let μ and ν be two measures on $X = \{x_1, x_2, x_3\}$. Let μ be additive with measures in the singletons equal to 0.2, 0.3, and 0.5. Let $\nu(\{x_1\}) = 0.2$, $\nu(\{x_1, x_2\}) = 0.2$, $\nu(\{x_1, x_3\}) = 0.2$, $\nu(X) = 1$ and $\nu(A) = 0$ for all the other sets. Then, there is no assignment that is consistent with the marginals of τ_μ and τ_ν. Table 1 displays the measures for all subsets of X, the values of τ_μ and τ_ν, and an assignment that is consistent with all marginals except the one for $\tau_\nu(X)$.

In order that there is a possible assignment for any pair of measures, we need to relax the problem. The relaxation is to allow lack of assignment for any set and measure. More formally, for a measure μ and the set A, we consider the value $lack_\mu(A)$, and for ν and the set B, we consider the value $lack_\nu(B)$. Then, an assignment $assg$ needs to be consistent with the measures given $lack_\mu$ and $lack_\nu$. In other words, the following two equations need to hold:

$$\mu(A) = \max_{A' \subset A} \mu(A') + \sum_{\emptyset \subset B' \subseteq X} assg(A, B') + lack_\mu(A)$$

$$\nu(B) = \max_{B' \subset B} \nu(B') + \sum_{\emptyset \subset A' \subseteq X} assg(A', B) + lack_\nu(B)$$

It is easy to see that these equations imply that the two rightmost terms need to correspond to the $(\max, +)$-transforms $\tau_\mu(A)$ and $\tau_\nu(B)$, respectively. We will use these equivalent equations using the transform in Definition 12. Table 2 gives a possible solution adapting the previous non-feasible assignment with the $lack$ values. More particularly, we define $lack_\nu(X) = -0.7$ to make the equality possible. Nevertheless, this solution has a negative value. We present a formalization below that forces all assignments to be non-negative. We will prove that there are feasible solutions in this case.

We formalize the optimal transport problem considering three cost functions. One associated to $assg$, called call c_a, and one associated to each measure, called c_μ and c_ν. These latter cost functions are naturally associated to $lack_\mu$ and $lack_\nu$, respectively.

Definition 12. Let μ and ν be two non-additive measures on the reference set X with $(\max, +)$-transform τ_μ and τ_ν, respectively. Then, the transport problem between μ and ν corresponds to find the functions $assg : (2^X \setminus \emptyset) \times (2^X \setminus \emptyset) \to [0, 1]$, $lack_\mu : (2^X \setminus \emptyset) \to [0, 1]$, and $lack_\nu : (2^X \setminus \emptyset) \to [0, 1]$ that satisfy

$$\tau_\mu(A) = \sum_{\emptyset \subset B' \subseteq X} assg(A, B') + lack_\mu(A)$$

$$\tau_\nu(B) = \sum_{\emptyset \subset A' \subseteq X} assg(A', B) + lack_\nu(B)$$

We will denote the solution of such problem by the tuple $(assg, lack_\mu, lack_\nu)$.

Then, given cost functions $c_a : 2^X \times 2^X \to [0, 1]$, $c_\mu : 2^X \to [0, 1]$, and $c_\nu : 2^X \to [0, 1]$, the cost of $(assg, lack_\mu, lack_\nu)$ is:

$$\sum_{\emptyset \subset A \subseteq X} \sum_{\emptyset \subset B \subseteq X} c_a(A, B) assg(A, B) + \sum_{A \subseteq X} c_\mu(A) lack_\mu(A) + \sum_{B \subseteq X} c_\nu(B) lack_\nu(B).$$

It is now possible to prove that for any pair of measures, there is at least an assignment that solves this transport problem.

Proposition 13. *Let μ and ν be two non-additive measures on the reference set X with $(\max, +)$-transform τ_μ and τ_ν, respectively. Then, there exist an assignment that solves the problem stated in Definition 12.*

Proof. Consider the assignment $lack_\mu(A) = \tau_\mu(A)$ and $lack_\nu(B) = \tau_\nu(B)$, and $assg(A, B) = 0$ for all other sets. This assignment satisfies the requirements in Definition 12. □

The following example gives an example of solution for Example 11.

Example 14. Let us reconsider the measures in Example 11. Table 3 shows a feasible solution for the transport problem according to Definition 12.

We have defined the problem in a way that $assg$ is not defined for empty sets. So, we can revisit the definition and express $lack_\mu(A)$ and $lack_\nu(B)$ as equal to $assg(A, \emptyset)$ and $assg(\emptyset, B)$, respectively. We can also proceed in the same way with the cost function extending it to any subset of X including the empty set. The cost of an assignment to the empty set is the cost associated to $lack_\mu$ and

Table 1
Measures μ and ν of Example 11, their (max, +)-transforms τ_μ and τ_ν, and a non-feasible assignment.

$\nu(B)$	τ_ν	set							
1	0.8	$\{x_1, x_2, x_3\}$	0	0.3	0.5	0.1	0.1	0.4	0.1
0	0	$\{x_2, x_3\}$	0	0	0	0	0	0	0
0.2	0	$\{x_1, x_3\}$	0	0	0	0	0	0	0
0.2	0	$\{x_1, x_2\}$	0	0	0	0	0	0	0
0	0	$\{x_3\}$	0	0	0	0	0	0	0
0	0	$\{x_2\}$	0	0	0	0	0	0	0
0.2	0.2	$\{x_1\}$	0.2	0	0	0	0	0	0
set			$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\{x_1, x_2, x_3\}$
τ_μ	–		0.2	0.3	0.5	0.1	0.1	0.4	0.1
$\mu(A)$			0.2	0.3	0.5	0.4	0.6	0.9	1.0

Table 2
Measures μ and ν of Example 11, their (max, +)-transforms τ_μ and τ_ν, and an assignment with $lack_\mu$ and $lack_\nu$. This example does not conform with Definition 12 because the assignment to $lack_\nu$ includes a negative value.

$\nu(B)$	τ_ν	set	$lack_\nu$							
1	0.8	$\{x_1, x_2, x_3\}$	-0.7	0	0.3	0.5	0.1	0.1	0.4	0.1
0	0	$\{x_2, x_3\}$	0	0	0	0	0	0	0	0
0.2	0	$\{x_1, x_3\}$	0	0	0	0	0	0	0	0
0.2	0	$\{x_1, x_2\}$	0	0	0	0	0	0	0	0
0	0	$\{x_3\}$	0	0	0	0	0	0	0	0
0	0	$\{x_2\}$	0	0	0	0	0	0	0	0
0.2	0.2	$\{x_1\}$	0	0.2	0	0	0	0	0	0
$lack_\mu$			0	0	0	0	0	0	0	0
set			–	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\{x_1, x_2, x_3\}$
τ_μ	–		–	0.2	0.3	0.5	0.1	0.1	0.4	0.1
$\mu(A)$			–	0.2	0.3	0.5	0.4	0.6	0.9	1.0

Table 3
Measures μ and ν of Example 11, their (max, +)-transforms τ_μ and τ_ν, and a feasible assignment with $lack_\mu$ and $lack_\nu$ according to Definition 12.

$\nu(B)$	τ_ν	set	$lack_\nu$							
1	0.8	$\{x_1, x_2, x_3\}$	0	0	0.3	0.5	0	0	0	0
0	0	$\{x_2, x_3\}$	0	0	0	0	0	0	0	0
0.2	0	$\{x_1, x_3\}$	0	0	0	0	0	0	0	0
0.2	0	$\{x_1, x_2\}$	0	0	0	0	0	0	0	0
0	0	$\{x_3\}$	0	0	0	0	0	0	0	0
0	0	$\{x_2\}$	0	0	0	0	0	0	0	0
0.2	0.2	$\{x_1\}$	0	0.2	0	0	0	0	0	0
$lack_\mu$			–	0	0	0	0.1	0.1	0.4	0.1
set			–	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\{x_1, x_2, x_3\}$
τ_μ	–		–	0.2	0.3	0.5	0.1	0.1	0.4	0.1
$\mu(A)$			–	0.2	0.3	0.5	0.4	0.6	0.9	1.0

$lack_\nu$. Using this approach, we can express the transport problem as follows. We also define below the optimal transport which is, of course, a solution with a minimum cost.

Definition 15. Let μ and ν be non-additive measures on X, with (max, +)-transforms τ_μ and τ_ν, respectively. Then, the transport problem between μ and ν is a function $assg : 2^X \times 2^X \to [0, 1]$ that satisfies

$$assg(\emptyset, \emptyset) = 0$$

$$\tau_\mu(A) = \sum_{B' \subseteq X} assg(A, B') \text{ for all } A \neq \emptyset$$

$$\tau_\nu(B) = \sum_{A' \subseteq X} assg(A', B) \text{ for all } B \neq \emptyset$$

Then, given the cost function $c_a : 2^X \times 2^X \to [0, 1]$, the cost of the assignment $assg$ is:

$$cost(c_a, assg) = \sum_{A \subseteq X} \sum_{B \subseteq X} c_a(A, B) assg(A, B).$$

Definition 16. Let μ, ν, τ_μ, τ_ν, and $assg$ as in Definition 15. Then, the optimal problem is to find an assignment $assg$ that minimizes $cost(c_a, assg)$.

Let us consider an example in which two additive measures are involved.

Example 17. Let μ and ν be two additive measures on X. The measure μ is defined by $\mu(\{x_1\}) = 0.2$, $\mu(\{x_2\}) = 0.3$, and $\mu(\{x_3\}) = 0.5$. The measure ν is defined by $\nu(\{x_1\}) = 0.2$, $\nu(\{x_2\}) = 0.2$, and $\nu(\{x_3\}) = 0.6$. Table 4 includes the measures as

Table 4
Additive case. This is not the only assignment. An alternative is $assg(\{x_2\},\{x_3\}) = 0.1$ instead of $assg(\{x_2\},\{x_2,x_3\}) = 0.1$ (and then replace $assg(\{x_1,x_2,x_3\},\{x_3\}) = 0.1$ by $assg(\{x_1,x_2,x_3\},\{x_2,x_3\}) = 0.1$).

$\nu(B)$	τ_ν	set								
1	0.2	$\{x_1,x_2,x_3\}$	0	0	0	0	0	0	0	0.2
0.8	0.2	$\{x_2,x_3\}$	0	0	0	0	0	0	0.2	0
0.8	0.2	$\{x_1,x_3\}$	0	0	0	0	0	0.2	0	0
0.4	0.2	$\{x_1,x_2\}$	0	0	0	0	0.2	0	0	0
0.6	0.6	$\{x_3\}$	0	0	0.1	0.5	0	0	0	0
0.2	0.2	$\{x_2\}$	0	0	0.2	0	0	0	0	0
0.2	0.2	$\{x_1\}$	0	0.2	0	0	0	0	0	0
0	0	$\emptyset$	0	0	0	0	0	0	0.1	0
set			$\emptyset$	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1,x_2\}$	$\{x_1,x_3\}$	$\{x_2,x_3\}$	$\{x_1,x_2,x_3\}$
τ_μ	–		0.0	0.2	0.3	0.5	0.2	0.2	0.3	0.2
$\mu(A)$			0.0	0.2	0.3	0.5	0.5	0.7	0.8	1.0

well as the $(\max,+)$-transforms τ_μ and τ_ν. A feasible assignment *assg* is also included.

4. Results

In this section we provide some results related to our definitions. In particular, we show that our definitions generalize standard optimal transport problems. That is, when the two measures are additive, the solution of our problem corresponds to the solution of a classical optimal transport problem. We also discuss how to solve the problems defined. For the sake of generality, we will consider pairs of measures μ and ν on reference sets X and Y.

4.1. Optimal transports as proper generalizations

One can ask how a solution for this problem in the setting of non-additive measures relates to the one we would obtain with the optimal transport problem in the classical probabilistic setting. This question can be stated in the following terms.

Let μ and ν be additive measures, is the optimal assignment according to the problems defined in this section the optimal assignment using the standard definition?

First, observe that one optimal assignment for probabilities is defined considering pairs (x,y) from (X,Y). In contrast, optimal assignments for non-additive measures are defined considering pairs (A,B) that are subsets of 2^X and 2^Y. Because of that, we consider the equivalence of the latter when restricted to the pairs considered by the former.

Definition 18. Let a be an assignment for pairs (x,y) where $x \in X$ and $y \in Y$ and let *assg* be an assignment for pairs (A,B) where $A \subseteq X$ and $B \subseteq Y$. Then, we define $a \prec assg$ if and only if $a(x,y) = assg(\{x\},\{y\})$ for all $x \in X$ and $y \in Y$.

It should be clear that there may be multiple (infinite) assignments *assg* for an assignment a. Then, we can prove the following for the optimal transport problem based on basic probability assignments based on Definition 6.

Proposition 19. *Let μ and ν be two probability measures on finite reference sets X and Y, let c be the cost function of the optimal transport for probability measures, and let κ be an arbitrary value such that $\kappa > \max c(x,y)$. Then, let us define $c_b(\{x\},\{y\}) = c(x,y)$ for all $x \in X$ and $y \in Y$, and $c_b(A,B) = \kappa$ for all other pairs (A,B). Finally, let a be the optimal transport for probability distributions μ and ν for c, and let assg be the optimal transport according to Definition 6.*

Then, the optimal assignments a and assg are such that $a \prec assg$, and the values of the objective functions of these two problems are equal.

Proof. It is easy to see that the basic probability assignment τ_μ and τ_ν for both measures μ and ν will have zero value in all elements that are not singletons. So, *assg* will only take values on singletons. As c is equal to c_b on these sets, the proposition is proven. □

Similarly, we can prove the following with respect to the optimal transport problem for arbitrary measures in terms of Möbius transforms. That is, using the objective function in Eq. (5).

Proposition 20. *Let μ, ν, c be as in Proposition 19. Let c_M be a cost function defined from c using the approach for c_b in Proposition 19. Let a the optimal transport for μ and ν using c, and assg be the optimal transport obtained using Eq. (5) using c_M. Then, $a \prec assg$, and the values of the objective functions of these two problems are equal.*

Proof. Note that while *assg* can take negative values, the objective function takes the absolute value of these assignments. Therefore, any solution with smaller assignment on the pairs (x,y) will need to have larger values for non-singletons and, thus, the objective function will be larger. □

Let us now consider the optimal transport with respect to the $(\max,+)$-transform. While for any probability measure the basic probability assignment and, in general, the Möbius transform is zero for non-singletons, this is not the case for the $(\max,+)$-transform. Nevertheless, we can also obtain a similar theorem.

Proposition 21. *Let μ and ν be two probability measures on finite reference sets X and Y, let c be the cost function of the optimal transport problem for probability measures. Let κ be an arbitrary value such that $\kappa > \max c(x,y)$.*

Then, let us define a cost function c_a as follows: $c_a(\{x\},\{y\}) = c(x,y)$ for all $x \in X$ and $y \in Y$, $c_a(\{x\},B) = c_a(A,\{y\}) = \kappa$ for non-singletons $A \subseteq X$ and $B \subseteq X$, and $c_a(A,B) = 0$ for all other pairs (A,B). Let a be the optimal transport for probability distributions for μ and ν using c, and let assg be the optimal transport according to Definition 15 using this cost function c_a.

Then, the optimal assignments a and assg are such that $a \prec assg$, and the objective functions of the two problems are the same.

Proof. As μ and ν are probability measures, the $(\max,+)$-transforms on the singletons will correspond to their probabilities. As $c = c_a$ for the singletons, the optimal assignment *assg* for the singletons is the optimal assignment a. Observe that as $c_a(\{x\},B)$ and $c_a(A,\{y\})$ is κ for non-singletons (or for $A = \emptyset$ and $B = \emptyset$), no other assignment *assg* for these sets can lead to a lower value. Then, there will some arbitrary assignments for the pairs (A,B) when both are non-singletons (including $A = \emptyset$ and $B = \emptyset$).

The contribution of singletons to the objective function will be equivalent to the objective function of a. The contribution of non-

Table 5
Additive case. Mass assigned to empty sets.

$\nu(B)$	τ_ν	set								
1	0.2	$\{y_1, y_2, y_3\}$	0.2	0	0	0	0	0	0	0
0.8	0.2	$\{y_2, y_3\}$	0.2	0	0	0	0	0	0	0
0.8	0.2	$\{y_1, y_3\}$	0.2	0	0	0	0	0	0	0
0.4	0.2	$\{y_1, y_2\}$	0.2	0	0	0	0	0	0	0
0.6	0.6	$\{y_3\}$	0.6	0	0	0	0	0	0	0
0.2	0.2	$\{y_2\}$	0.2	0	0	0	0	0	0	0
0.2	0.2	$\{y_1\}$	0.2	0	0	0	0	0	0	0
0	0	$\emptyset$	0	0.2	0.3	0.5	0.2	0.2	0	0
set			$\emptyset$	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$	$\{x_1, x_2, x_3\}$
τ_μ	–		0.0	0.2	0.3	0.5	0.2	0.2	0.3	0.2
$\mu(A)$			0.0	0.2	0.3	0.5	0.5	0.7	0.8	1.0

singletons to the objective function will be zero. Therefore, the proposition is proven. □

The properties in this section mean that for any cost function c and probability measures μ and ν, we have a cost function c_a such that the optimal transport problem for c corresponds to the optimal transport problem using basic probability assignments for belief functions, Möbius transform for arbitrary non-additive measures, and $(\max, +)$-transform also for an arbitrary non-additive measure.

In terms of the optimal transport problem, these propositions mean that our definition is a generalization of the classical one in the sense that if we apply our approach to additive measures we can reproduce the same solutions we would obtain with the classical optimal transport problem. In addition, with our approach we can solve the case of non-additive measures.

The propositions establish that the assignment on the singletons will be the same. Nevertheless, nothing is said on the assignment for the other sets. Observe that for both basic probability assignments and Möbius transforms the mass for non-singletons will be zero, so, the optimal assignment for our problems will be also zero for these sets. In contrast, the $(\max, +)$-transform is not zero for the non-singletons and, therefore, the assignment will neither be zero for them.

4.2. Some additional properties

The optimal transport problem according to Definition 15 has an interesting property. Given two measures, it is possible to *transfer* all the mass from sets $\emptyset \neq A \subseteq X$ to the empty set (i.e., what corresponds to the lack variable $lack_\mu(A)$) and the same for all sets $\emptyset \neq B \subseteq Y$. This would produce the solution described in Table 5. To avoid this type of transfer we need that the cost from a set A to the emptyset, and from emptyset to a set B should be larger than a direct assignment from A to B. That is,

$$c_a(A, \emptyset) + c_a(\emptyset, B) \geq c_a(A, B).$$

In Proposition 21 we have given a cost function c_a that provides a solution that is compatible with the optimal transport problem for probabilities. We can have other assignments that have the same property. Given that $c(x, y) < \infty$, we can use a cost function c_a satisfying the following conditions:

- $c_a(\{x\}, \{y\}) = c(x, y)$,
- $c_a(\{x\}, \emptyset) = \kappa^+ > \max_{y_i \in Y} c(\{x\}, \{y_i\})$,
- $c_a(\emptyset, \{y\}) = \kappa^+ > \max_{x_i \in X} c(\{x_i\}, \{y\})$,
- $c_a(A, \{y\}) = \kappa > \max_{x_i \in A} c(\{x_i\}, \{y\})$ for $|A| > 1$, and
- $c_a(\{x\}, B) = \kappa > \max_{y_i \in B} c(\{x\}, \{y_i\})$ for $|B| > 1$.
- $c_a(A, B) = 0$ for $|A| > 1$ and $|B| > 1$ to satisfy Proposition 21. Here we assume $\kappa^+ > \kappa$.

Table 6
Graphical representation of a cost function c_a so that its corresponding optimal assignment *assg* is compatible with another assignment a for the cost function c. That is, $a \prec assg$.

Non-singletons $\lvert B\rvert > 1$	κ^+	κ	†
singletons $\{y\}$	κ^+	c	κ
$\emptyset$	–	κ^+	κ^+
–	$\emptyset$	singletons $\{x\}$	non-singletons $\lvert A\rvert > 1$

This structure, depicted in Table 6, avoids any assignment from a singleton to a larger set as the cost becomes κ, and, thus, prioritizes the assignment between singletons, even in the case that μ and ν are not probabilities.

This type of assignment can be represented as a graph with two components (i.e., two sets of nodes, and then the nodes in a set cannot be reached from nodes in the other set). One set corresponds to the singletons and the other to the non-singletons. When $|X| = |Y|$ we can define cost functions where this property holds for each cardinality (or for some cardinalities). In this case, we can also represent the structure in terms of graphs where we have a component for each cardinality.

The cost among non-singletons is zero so that the two assignments a and *assg* have the same overall costs. Other costs different to zero will provide the same assignments a and *assg* but the overall costs may be different.

Proposition 21 proves that given c and probabilities μ and ν we have an equivalent problem (c_a, μ, ν) that returns the same assignment for the singletons. We can observe that each of the values $p(x) = \mu(\{x\})$ (same for $q(y) = \nu(\{y\})$) appears a different number of times. So, we may consider extending c into c_a so that the cost of the assignments is similar. In other words, for a pair $(p(x_i), q(y_j))$, the cost of all related assignments in c_a is the same as $c(x_i, y_j)$. For example, if we consider the pair $(p(x_1) = 0.2, q(y_3) = 0.6)$ we have that this pair appears 4×1 times in Table 5. This observation allows us to construct a cost function for our optimal transport problem in which the cost associated to this pair is $(p(x_1) = 0.2, p(y_3) = 0.6)$, which is the same as the one of $c(x_1, x_3)$.

Proposition 22. *Let μ and ν be additive measures on X and Y and let p and q be the corresponding probability distributions. Let c be the cost function of the optimal transport problem in the probabilistic setting associated to p and q. Then, there exists a cost function c_a such that the cost of each pair of probabilities $(p(x_i), q(y_j))$ equals to $c(x_i, y_j)$.*

The proof of this proposition is by construction, and the expression for the cost function c_a is given in Eq. (7)

Proof. First, note that $\tau_\mu(A) = \min_{x\in A} p(x)$ and $\tau_\nu(A) = \min_{x\in A} q(x)$ according to Proposition 4. It is obvious that for singletons $\{x_i\}$ and $\{y_j\}$ the value $c_a(\{x_i\},\{y_j\})$ should be associated to $c(x_i, y_j)$. Similarly, in general, we need to associate $c_a(A, B)$ to the pair (x, y) such that it produces $\tau_\mu(A)$ and $\tau_\nu(B)$. In other words, we associate $c_a(A, B)$ to $c(x_i, x_j)$ where $x_i = \arg\min_{x_i\in A} p(x_i)$ and $y_j = \arg\min_{y_j\in B} q(y_j)$.

The association between c_a and c is not, in general, an equality. Some probabilities p_i appear several times in $\tau_\mu(A)$. The same applies to q_i. In fact, the smallest probability $\min_{x_i} p(x_i)$ appears more often than the second smallest one, which appears more often than the third smallest, and so on.

Let $s_\mu(i)$ and $s_\nu(i)$ define a permutation so that

$$p\left(x_{s_\mu(1)}\right) \le p\left(x_{s_\mu(2)}\right) \le \ldots \le p\left(x_{s_\mu(n)}\right) \tag{6}$$

$$q\left(x_{s_\nu(1)}\right) \le q\left(x_{s_\nu(2)}\right) \le \ldots \le q\left(x_{s_\nu(m)}\right)$$

Then, $p(x_{s_\mu(n)})$ appears in $\tau_{\mu(A)}$ only in one set (i.e., $\{x_{s_\mu(n)}\}$), the second largest value $p(x_{s_\mu(n-1)})$ appears in two sets (i.e., $\{x_{s_\mu(n-1)}\}$, $\{x_{s_\mu(n)}, x_{s_\mu(n-1)}\}$). In general, the ith largest value appears in 2^{n-1} sets. The same applies to the values for q. Observe Table 4. For τ_μ we have that $p(\{x_3\}) = 0.5$ appears only once, $p(\{x_2\}) = 0.3$ appears twice, and $p(\{x_1\}) = 0.2$ appears 4 times. Same applies to q although in that case as $q(\{x_1\}) = q(\{x_2\}) = 0.2$, this is not so easily observed.

Let $s_\mu^{-1}(i)$ be the order of the ith variable, according to the permutation in Eq. (6). Similarly, $s_\nu^{-1}(j)$ corresponds to the order of the jth variable. Then, if the pair associated to (A, B) is (x_i, x_j), we need to weight the associated cost $c(x_i, x_j)$ considering the fact that $p(x_i)$ probability appears several times. I.e., $p(x_i)$ appears $2^{n-s_\mu^{-1}(i)}$ times. Similarly, $q(y_j)$ appears $2^{m-s_\nu^{-1}(j)}$ times. Therefore, we define:

$$\begin{aligned} c_a(A, B) &= \frac{1}{2^{n-i'}}\frac{1}{2^{m-j'}} c\left(\arg\min_{x_i\in A} p(x_i), \arg\min_{y_j\in B} q(y_j)\right) \\ &= \frac{1}{2^{n-s_\mu^{-1}(i)}}\frac{1}{2^{m-s_\nu^{-1}(j)}} c(x_i, y_j) \end{aligned} \tag{7}$$

where $i = \arg\min_{x_i\in A} p(x_i)$ and $j = \arg\min_{y_j\in B} q(y_j)$.

Then, we can observe that the pair $(p(x_{s_\mu(n)}), q(y_{s_\mu(m)}))$ appears only once in both τ_μ and τ_ν and

$$c_a(\{x_{s_\mu(n)}\},\{y_{s_\mu(m)}\}) = c(x_{s_\mu(n)}, y_{s_\mu(m)})$$

Then, for a pair $\{x_i\},\{y_j\}$ we have that $p(x_i)$ will appear $2^{n-s_\mu^{-1}(i)}$ times and $q(y_j)$ will appear $2^{n-s_\nu^{-1}(j)}$ times. I.e., there will be $2^{n-s_\mu^{-1}(i)}2^{m-s_\nu^{-1}(j)}$ pairs of sets (A, B) with c_a associated to the same c. □

If μ and ν are additive measures; then, if we define c_a according to Eq. (7), and we compute the solutions of the original problem (i.e., a) and of our extension (i.e., *assg*), then they are such that $a \prec assg$. In addition the two assignments have the same cost. This is what this propositions states.

Nevertheless, this result is about building the cost function c_a for non-singletons taking into account the order of the values for the singletons. That is, using the ordering or permutation of elements $s_\mu(1), \ldots, s_\mu(n)$, as well as $s_\nu(1), \ldots, s_\nu(n)$. That is, the cost function c_a depends on both μ and ν. In other words, we cannot define a cost function c_a a priori and use it to evaluate an arbitrary pair of measures μ and ν.

The next definition defines this cost function explicitly. It follows Eq. (7). We write $c_a(A, B; \mu, \nu)$ to make this dependence on μ and ν explicit. Note also that this definition is for arbitrary fuzzy measures and not only for additive ones. Then, we provide the proposition that establishes the consistency of the assignments.

Definition 23. Let μ and ν be non-additive measures on X and Y. Let κ be an arbitrary value such that $\kappa > \max c(x, y)$. Then, we define the cost function $c_a : 2^X \times 2^Y \to \mathbb{R}^+$ as follows

$$c_a(A, B; \mu, \nu) = \begin{cases} 0 & A = \emptyset, B = \emptyset \\ \kappa & A = \emptyset, B \neq \emptyset \\ \kappa & A \neq \emptyset, B = \emptyset \\ \frac{1}{2^{n-s_\mu^{-1}(i)}}\frac{1}{2^{m-s_\nu^{-1}(j)}} c(x_i, y_j) & \text{otherwise} \end{cases}$$

where n is the cardinality of X, m is the cardinality of Y, and s_μ and s_ν are defined as above in Eq. (6).

Proposition 24. *Let μ and ν be additive measures on X and Y. Let c be a cost function on $X \times Y$, and let c_a be the cost function according to Definition 23. Let a be the solution of the optimal transport problem for c, and let assg be the solution of the optimal transport problem for c_a. Then, a and assg are such that $a \prec assg$.*

Proof. This is a consequence of Proposition 22 as the cost of each pair in c_a is the same as in the cost function c. In addition, as $c_a(\emptyset, B) = c_a(A, \emptyset) > c(\{x\},\{y\})$ for all x, y, then there is no assignment to $\emptyset$. □

4.3. On the implementation of the optimal transport problem

We discuss in this section how to find a solution for the optimal transport problem for the three different problems established above. We will start with the case of basic probability assignment.

It can be observed that the problem introduced in Definition 6 can be seen as an assignment of probabilities in 2^X instead of probabilities on X. Therefore, the same techniques and approaches used for classical optimal transport problems can be applied here.

In contrast, the problem stated in Eq. (5) needs a different approach. We can observe that it has some similarities with the one for basic probability assignment but in this case the values of the Möbius transform can be negative. In addition, the objective function includes an absolute value function. An objective function with an absolute value of a linear expression can be transformed into an equivalent one without the absolute value by means of introducing a new variable. The problem will still be linear. See e.g., Granger et al. (2022). More particularly, if we have an objective function of this form

$$\min OF + |t|$$

we will rewrite it as

$$\min OF + t'$$

and add two additional constraints, $+t \le t'$ and $-t \le t'$. This process requires quite a few additional constraints (two for each variable). As there are $2^{|X|} \times 2^{|X|}$ variables, this corresponds to $2 \cdot 2^{2|X|}$ additional constraints.

The optimal transport problem associated to the $(\max, +)$-transform has a form similar to the one for basic probability assignments. The assignment needs to be positive. Therefore, the problem is also a linear optimization problem with linear constraints. We have developed software in Python for the computation of this optimal problem. Software is available at[1].

4.4. Wasserstein distance

The definition of the optimal transport problem for non-additive measures permits us to consider the definition of distances based on this problem. More particularly, we can consider

[1] http://www.mdai.cat/code.

the definition of a Wasserstein-like discrepancy following Eq. (1). Our definition is based on the definition of the transport problem based on the $(\max, +)$-transform.

Definition 25. Let μ and ν be non-additive measures on X, with $(\max, +)$-transforms τ_μ and τ_ν, respectively. Let c_a be a cost function. Let $\Pi(\tau_\mu, \tau_\nu)$ be the set of all assignments that are compatible with τ_μ and τ_ν. Then, we define the Wasserstein-like discrepancy for μ and ν given c_a as

$$d_{c_a}(\mu, \nu) = \inf_{assg \in \Pi(\tau_\mu, \tau_\nu)} \sum_{A \subseteq X} \sum_{B \subseteq X} c_a(A, B) assg(A, B) \tag{8}$$

It is clear that when $\mu = \nu$ we have that $d_{c_a}(\mu, \nu)$ is zero, and that if there is a single optimal assignment, the expression is symmetric. Other properties of this definition need to be further studied. In particular, the triangle inequality so that the discrepancy is, in fact, a distance. Naturally, some properties will depend on the cost function, and on the type of solution the optimization problem gives as result.

5. Conclusions and future work

In this paper we have studied the transport problem for non-additive measures. We have discussed the difficulties of this problem. Then, we have proposed three different definitions, one for belief functions and basic probability assignments and two others for arbitrary non-additive measures. The latter are based on the Möbius transform and the $(\max, +)$-transform. Different formulations stress different ways of looking to the measures. They deal differently with positive and negative interactions. The Möbius transform can be negative, but not the $(\max, +)$-transform. They also deal differently with the fact that the total mass of a non-additive measure can be seen as different to one.

We have proven some properties, including the ones that establish that our definitions are proper extensions of standard optimal transport problems. We have also briefly discussed the computational complexity of the related optimization problems.

In the paper we have discussed the cost functions that we need to define on subsets of X and Y and some relationships with the ones for probability distributions. What kind of cost functions are relevant and meaningful for non-additive measures is a research direction. It is known that in the classical setting some cost functions ensure that the Wasserstein distance satisfies the triangle inequality and, thus, it is a proper distance. A research direction is to study this problem.

We have also provided a discussion on how the problems can be solved numerically. All the problems stated here are linear optimization problems with linear constraints. Nevertheless, the assignment is a function from $2^{|X|} \times 2^{|Y|}$ that makes the problem costly for not so large X and Y. Therefore another direction is to work on computational efficient solutions, some may depend on the type of cost functions used.

Acknowledgments

This work was partially supported by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

References

Abril, D., Navarro-Arribas, G., & Torra, V. (2012). Choquet integral for record linkage. *Annals of Operations Research, 195*, 97–110.
Agahi, H. (2020). A generalized Hellinger distance for Choquet integral. *Fuzzy Sets and Systems, 396*, 42–50.
Agahi, H., & Yadollahzadeh, M. (2021). On f-divergence for $\sigma - \oplus$-measures. *Soft Computing, 25*, 9781–9787.
Ambrosio, L., Gigli, N., & Savaré, G. (2008). *Gradient flows in metric spaces and in the space of probability measures*. Birkhäuser.
Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In *ICML* (pp. 214–223). PMLR.
Bardozzo, F., de la Osa, B., Horanská, L., Fumanal-Idocin, J., delli Priscoli, M., Troiano, L., … Bustince, H. (2021). Sugeno integral generalization applied to improve adaptive image binarization. *Information Fusion, 68*, 37–45.
Beliakov, B. (2022). Knapsack problems with dependencies through non-additive measures and Choquet integral. *European Journal of Operational Research, 301*(1), 277–286.
Benvenuti, P., Mesiar, R., Vivona, D., & Pap, E. (2002). Monotone set functions-based integrals. In *Handbook of measure theory* (pp. 1329–1379). North-Holland.
Bogachev, V. I., & Kolesnikov, A. V. (2012). The Monge–Kantorovich problem: Achievements, connections, and perspectives. *Russian Mathematical Surveys, 67*(5), 1–110.
Bronevich, A. G., & Rozenberg, I. N. (2021). The measurement of relations on belief functions based on the Kantorovich problem and the Wasserstein metric. *International Journal of Approximate Reasoning, 131*, 108–135.
Chateauneuf, A. (1996). Decomposable measures, distorted probabilities and concave capacities. *Mathematical Social Sciences, 31*, 19–37.
Chateauneuf, A., & Jaffray, J. Y. (1989). Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. *Mathematical Social Sciences, 17*(3), 263–283.
Choquet, G. (1953/54). Theory of capacities. *Annales de l'Institut Fourier, 5*, 131–295.
Denneberg, D. (1994). *Non additive measure and integral*. Kluwer Academic Publishers.
Dimuro, G. P., Lucca, G., Bedregal, B. R. C., Mesiar, R., Sanz, J. A., Lin, C.-T., & Bustince, H. (2020a). Generalized CF1F2-integrals: From Choquet-like aggregation to ordered directionally monotone functions. *Fuzzy Sets Systems, 378*, 44–67.
Dimuro, G. P., Fernández, J., Bedregal, B. R. C., Mesiar, R., Sanz, J. A., Lucca, G., & Bustince, H. (2020b). The state-of-art of the generalizations of the Choquet integral: From aggregation and pre-aggregation to ordered directionally monotone functions. *Information Fusion, 57*, 27–43.
Edwards, W. (1953). Probability-preferences in gambling. *American Journal of Psychology, 66*, 349–364.
Gal, S. G., & Niculescu, C. P. (2019). Kantorovich's mass transport problem for capacities. arXiv:1907.03749v4
Gilboa, I. (2009). *Theory of decision under uncertainty*. Cambridge University Press.
Granger, B., Yu, M., & Zhou, K. (2022). Optimization with absolute values. https://optimization.mccormick.northwestern.edu/index.php/Optimization_with_absolute_values(accessed 9 March 2022).
Honda, A., Nakano, T., & Okazaki, Y. (2002a). Distortion of fuzzy measures. In *Proc. of the SCIS/ISIS conference*.
Honda, A., Nakano, T., & Okazaki, Y. (2002b). Subjective evaluation based on distorted probability. In *Proc. of the SCIS/ISIS conference*.
Honda, A., & Narukawa, M. S. (2014). Entropy of capacity. In Y. V. Torra (Ed.), *Non-additive measures* (pp. 79–95). Springer.
Huizing, G. J., Peyré, G., & Cantini, L. (2022). Optimal transport improves cell-cell similarity inference in single-cell omics data. *Bioinformatics, 38*, 2169–2177.
Kantorovich, L. V. (1942). On mass moving. *Doklady Akademii Nauk SSSR, 37*(7-8), 227–229.
Marco-Detchart, C., Lucca, G., Lopez-Molina, C., De Miguel, L., Pereira Dimuro, G., & Bustince, H. (2021). Neuro-inspired edge feature fusion using Choquet integrals. *Information Sciences, 581*, 740–754.
Mayag, B., & Bouyssou, D. (2020). Necessary and possible interaction between criteria in a 2-additive Choquet integral model. *European Journal of Operational Research, 283*, 308–320.
Mesiar, R. (1997). k-order pan-discrete fuzzy measures. In *Proc. 7th IFSA world congress* (pp. 488–490).
Mesiar, R. (1999). Generalizations of k-order additive discrete fuzzy measures. *Fuzzy Sets and Systems, 102*, 423–428.
Mesiar, R., Kolesárová, A., Bustince, H., Dimuro, G. P., & Bedregal, B. C. (2016). Fusion functions based discrete Choquet-like integrals. *European Journal of Operational Research, 252*(2), 601–609.
Mesiar, R., & Mesiarová, A. (2004). Fuzzy integrals. In *MDAI 2004*. In *Lecture notes in artificial intelligence: vol. 3131* (pp. 7–14).
Narukawa, Y., Torra, V., & Sugeno, M. (2016). Choquet integral with respect to a symmetric fuzzy measure of a function on the real line. *Annals of Operations Research, 244*(2), 571–581.
Ontkovičová, Z., & Kiselák, J. (2022). A way to proper generalization of ϕ-divergence based on Choquet derivatives. *Soft Computing, 26*, 11295–11314.
Pereira, M. A., Figueira, J. R., & Marques, R. C. (2020). Using a Choquet integral-based approach for incorporating decision-makers preference judgments in a data envelopment analysis model. *European Journal of Operational Research, 284*(3), 1016–1030.
Santambrogio, F. (2015). Optimal transport for applied mathematicians - calculus of variations. In *PDEs and modeling, Birkhäuser*.
Sugeno, M. (1972). Fuzzy measures and fuzzy integrals (in Japanese). *Transactions of the Society of Instrument and Control Engineers, 8*(2).
Sugeno, M. (1974). *Theory of fuzzy integrals and its applications*. Tokyo, Japan: Tokyo Institute of Technology Ph.D. dissertation.
Tehrani, A. F., Cheng, W., Dembczynski, K., & Hüllermeier, E. (2012). Learning monotone nonlinear models using the Choquet integral. *Machine Learning, 89*, 183–211.

Torra, V. (2017). Entropy for non-additive measures in continuous domains. *Fuzzy Sets and Systems, 324*, 49–59.
Torra, V. (2022). $(\max, \oplus)$-transforms and genetic algorithms for fuzzy measure identification. *Fuzzy Sets and Systems, 451*, 253–265. https://doi.org/10.1016/j.fss.2022.09.008.
Torra, V., Guillen, M., & Santolino, M. (2018). Continuous m-dimensional distorted probabilities. *Information Fusion, 44*, 97–102.
Torra, V., Narukawa, Y., & Abril, D. (2014). Comparing fuzzy measures through their Möbius transform. In *Proc. 17th int. conf. information fusion (FUSION)*.
Torra, V., Narukawa, Y., & Sugeno, M. (2013a). *Non-additive measures: Theory and applications*. Springer.
Torra, V., Narukawa, Y., Sugeno, M., & Carlson, M. (2013b). Hellinger distance for fuzzy measures. In *EUSFLAT 2013*.
Torra, V., Narukawa, Y., & Sugeno, M. (2016). On the f-divergence for non-additive measures. *Fuzzy Sets and Systems, 292*, 364–379.
Torra, V., Narukawa, Y., & Sugeno, M. (2020). On the f-divergence for discrete non-additive measures. *Information Sciences, 512*, 50–63.
Villani, C. (2003). *Topics in optimal transportation*. AMS.
Villani, C. (2008). *Optimal Transport: Old and New*. Springer.