Optimal Transport in Dempster-Shafer Theory and Choquet-Wasserstein Pseudo-Distances

Silvia Lorenzini[1][0009−0008−1785−3065], Davide Petturiti[1][0000−0002−3277−4217], and Barbara Vantaggi[2][0000−0002−3651−3743]

[1] Dept. Economics, University of Perugia, Perugia, Italy
`silvia.lorenzini@dottorandi.unipg.it`
`davide.petturiti@unipg.it`
[2] Dept. MEMOTEF, Sapienza University of Rome, Rome, Italy
`barbara.vantaggi@uniroma1.it`

Abstract. We consider the marginal problem in Dempster-Shafer theory, investigating the structure of a suitable set of bivariate joint belief functions having fixed marginals, by relying on copula theory. Next, we formulate two Kantorovich-like optimal transport problems, either seeking to minimize the Choquet integral of a given cost function with respect to the reference set of joint belief functions or its dual functional. We finally give a noticeable application by choosing a metric as cost function: this permits to define pessimistic and optimistic Choquet-Wasserstein pseudo-distances, that can be used to compare belief functions on the same space.

Keywords: Belief functions · Optimal transport · Choquet-Wasserstein pseudo-distances.

1 Introduction

The *optimal transport (OT)* problem has gathered an increasing attention in the probabilistic setting during the last decades, due to its numerous connections with other branches of mathematics [19] and its plethora of applications in computer vision, computer graphics, image processing, statistics and machine learning [14]. As is well-known, OT relies on the probabilistic *marginal problem* and its popularity is also connected to the *Wasserstein distance* [20], that provides a metric on the set of probability measures on a metric space. In turn, the widespread use of the Wasserstein distance in machine learning techniques like WGANs [1] has been favored by *entropic regularizations* [13], that allow to design efficient optimization algorithms.

The idea of OT naturally extends to the case where the marginal distributions convey ambiguity in the sense of [6]. In such cases, the Dempster-Shafer theory [4, 16] reveals to be the natural framework to encode uncertainty, departing the less from probability theory. This suggests to formulate OT in Dempster-Shafer theory, for which several definitions can be given: two possible approaches are those in [3] and [18].

In this work we start by considering the marginal problem in Dempster-Shafer theory and provide an analysis based on copula theory [12] of a suitable subset of bivariate joint belief functions with given marginals, introduced in [8, 17]. Next, restricting to such a set of joint belief functions, we formulate a pessimistic and an optimistic Kantorovich-like optimal transport problem, by minimizing the Choquet integral of a cost function or its dual functional, respectively. Such an approach differs from [3], where the OT is expressed in terms of Möbius inverse and a metric between sets, while in [18] the author refers to $(\max, +)$-transforms.

After that, choosing a metric as cost function, we define a pessimistic and an optimistic *Choquet-Wasserstein pseudo-distance* (where the term "pseudo-distance" is not given any metric connotation), proving that they are a dissimilarity function and a metric-like function (see [5]) on the space of belief functions, respectively. Finally, we introduce a suitable entropic regularization to find a "closest" *at most k-additive* belief function [9, 10] of a given belief function, according to the two pseudo-distances. Such a "closest" belief function can be practically computed by relying on an adaptation of *Dykstra's algorithm* [2, 13], that we omit due to space limitations.

The paper is structured as follows. Section 2 recalls the basic notions of Dempster-Shafer theory and Choquet integration. Section 3 presents the marginal problem in Dempster-Shafer theory and characterizes the reference subset of joint belief functions with fixed marginals. Section 4 introduces the pessimistic and optimistic optimal transport problems in Dempster-Shafer theory. Finally, Section 5 defines the pessimistic and optimistic Choquet-Wasserstein pseudo-distances and show their use in finding a "closest" at most k-additive belief function of a given belief function, while Section 6 draws our conclusions. Proofs are not reported due to the limited number of pages.

2 A Glimpse of Dempster-Shafer Theory

Let $\Omega = \{\omega_1, \ldots, \omega_d\}$ be a finite non-empty set of states of the world endowed with the power set 2^Ω, and denote by $\mathbb{R}^\Omega$ the set of all random variables. A *belief function* (see [4, 16]) is a set function $\nu : 2^\Omega \to [0, 1]$ satisfying:

(i) $\nu(\emptyset) = 0$ and $\nu(\Omega) = 1$;

(ii) $\nu\left(\bigcup_{i=1}^{k} E_i\right) \geq \sum_{\emptyset \neq I \subseteq \{1,\ldots,k\}} (-1)^{|I|+1} \nu\left(\bigcap_{i \in I} E_i\right)$, for all $k \geq 2$, $\{E_i\}_{i=1}^k \subseteq 2^\Omega$.

A belief function ν is associated with a dual set function $\overline{\nu}$ on 2^Ω called *plausibility function* and defined, for all $A \in 2^\Omega$, as $\overline{\nu}(A) = 1 - \nu(A^c)$.

Every belief function ν, and so its dual plausibility function $\overline{\nu}$, is completely characterized by its *Möbius inverse* [10], that is a set function $m_\nu : 2^\Omega \to [0, 1]$ such that $m_\nu(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m_\nu(A) = 1$.

An event with strictly positive Möbius inverse is called a *focal element* and we denote by $\mathcal{F}_\nu = \{E \in 2^\Omega : m_\nu(E) > 0\}$ the set of focal elements of ν. We have that, for all $A \in 2^{\Omega}$, it holds:

$$\nu(A) = \sum_{B \in \mathcal{F}_\nu,\, B \subseteq A} m_\nu(B) \quad \text{and} \quad \overline{\nu}(A) = \sum_{B \in \mathcal{F}_\nu,\, B \cap A \neq \emptyset} m_\nu(B).$$

Looking at the set of focal elements $\mathcal{F}_\nu$, a belief function ν is called:

- *vacuous at* A, for $A \in 2^{\Omega}$, if $\mathcal{F}_\nu = \{A\}$ and in this case ν is denoted by δ_A;
- k-*additive*, for $1 \leq k \leq d$, if $|E| \leq k$, for all $E \in \mathcal{F}_\nu$, with at least an equality;
- *probability measure*, if it is 1-additive and in this case ν is denoted by π.

Given a belief function ν and $X \in \mathbb{R}^{\Omega}$, the *Choquet expectation* of X with respect to ν is defined through the Choquet integral [10]

$$\oint X \mathrm{d}\nu = \sum_{i=1}^{d} [X(\omega_{\sigma(i)}) - X(\omega_{\sigma(i+1)})]\nu(E_i^{\sigma}),$$

where σ is a permutation of $\{1, \ldots, d\}$ such that $X(\omega_{\sigma(1)}) \geq \cdots \geq X(\omega_{\sigma(d)})$, $E_i^{\sigma} = \{\omega_{\sigma(1)}, \ldots, \omega_{\sigma(i)}\}$, for $i = 1, \ldots, d$, and $X(\omega_{\sigma(d+1)}) = 0$. If ν reduces to a probability measure π, we have that $\oint X \mathrm{d}\pi = \int X \mathrm{d}\pi$, where the latter denotes a classical Stieltjes integral. On the other hand, the Choquet expectation with respect to a plausibility function $\overline{\nu}$ can be defined through duality as

$$\oint X \mathrm{d}\overline{\nu} = -\oint (-X) \mathrm{d}\nu.$$

We also have that every belief function ν is in one-to-one correspondence with the (closed and convex) set of probability measures dominating it, called *core*, and denoted by $\mathrm{core}(\nu)$. Finally, we recall that both Choquet expectations in Dempster-Shafer theory can be expressed either in terms of $\mathrm{core}(\nu)$ or m_ν (see, e.g., [10]): for all $X \in \mathbb{R}^{\Omega}$

$$\oint X \mathrm{d}\nu = \min_{\pi \in \mathrm{core}(\nu)} \int X \mathrm{d}\pi = \sum_{B \in \mathcal{F}_\nu} \left(\min_{\omega \in B} X(\omega) \right) m_\nu(B), \tag{1}$$

$$\oint X \mathrm{d}\overline{\nu} = \max_{\pi \in \mathrm{core}(\nu)} \int X \mathrm{d}\pi = \sum_{B \in \mathcal{F}_\nu} \left(\max_{\omega \in B} X(\omega) \right) m_\nu(B). \tag{2}$$

3 Marginal Problem in Dempster-Shafer Theory

Let $\mathcal{X} = \{x_1, \ldots, x_m\}$ and $\mathcal{Y} = \{y_1, \ldots, y_n\}$ be two finite sets endowed with the algebras $2^{\mathcal{X}}$ and $2^{\mathcal{Y}}$, and let μ and ν be two belief functions on $2^{\mathcal{X}}$ and $2^{\mathcal{Y}}$. We consider the product space $(\mathcal{X} \times \mathcal{Y}, 2^{\mathcal{X} \times \mathcal{Y}})$ and denote $\widetilde{2^{\mathcal{X}}} = \{A \times \mathcal{Y} : A \in 2^{\mathcal{X}}\}$ and $\widetilde{2^{\mathcal{Y}}} = \{\mathcal{X} \times B : B \in 2^{\mathcal{Y}}\}$, which are two sub-algebras of $2^{\mathcal{X} \times \mathcal{Y}}$ isomorphic to $2^{\mathcal{X}}$ and $2^{\mathcal{Y}}$, respectively. As usual, the sets $\mathcal{X}$ and $\mathcal{Y}$ can be interpreted as the ranges of two random variables X and Y, the latter being identified with the canonical projections $X(x, y) = x$ and $Y(x, y) = y$.

The *marginal problem* in Dempster-Shafer theory consists in finding a joint belief function $\gamma : 2^{\mathcal{X}\times\mathcal{Y}} \to [0,1]$ such that

$$\begin{cases} \gamma(A \times \mathcal{Y}) = \mu(A), \text{ for all } A \in 2^{\mathcal{X}}, \\ \gamma(\mathcal{X} \times B) = \nu(B), \text{ for all } B \in 2^{\mathcal{Y}}, \end{cases} \tag{3}$$

which is equivalent to finding a Möbius inverse $m_\gamma : 2^{\mathcal{X}\times\mathcal{Y}} \to [0,1]$ such that

$$\begin{cases} \displaystyle\sum_{D\subseteq A\times\mathcal{Y}} m_\gamma(D) = \mu(A), \text{ for all } A \in 2^{\mathcal{X}}, \\ \displaystyle\sum_{D\subseteq \mathcal{X}\times B} m_\gamma(D) = \nu(B), \text{ for all } B \in 2^{\mathcal{Y}}. \end{cases} \tag{4}$$

The set of solutions of (3) or, equivalently, (4) is denoted by

$$\mathbf{B}(\mu,\nu) = \left\{\gamma : 2^{\mathcal{X}\times\mathcal{Y}} \to [0,1] \,:\, \gamma \text{ is a belief function}, \gamma_{|\widetilde{2^{\mathcal{X}}}} = \mu, \gamma_{|\widetilde{2^{\mathcal{Y}}}} = \nu\right\}, \tag{5}$$

and is easily shown to be a closed and convex subset of $[0,1]^{2^{\mathcal{X}\times\mathcal{Y}}}$ endowed with the product topolgy, whose non-emptyness has been shown in [17]. In particular, since $\mathbf{B}(\mu,\nu)$ is obtained by solving a finite system of linear equalities and inequalities, its set of extreme points $\mathrm{ext}(\mathbf{B}(\mu,\nu))$ is finite.

In general, for $\gamma \in \mathbf{B}(\mu,\nu)$, it is not possible to know a priori the set of focal elements $\mathcal{F}_\gamma$ and this is the main difficulty in providing an explicit characterization of $\mathrm{ext}(\mathbf{B}(\mu,\nu))$ since $2^{m\cdot n} - 1$ variables must be considered to reconstruct m_γ.

Following [8] (see also [3, 11]) a more manageable subset of $\mathbf{B}(\mu,\nu)$ can be obtained by referring to the sets of focal elements $\mathcal{F}_\mu = \{E_1,\dots,E_M\}$ and $\mathcal{F}_\nu = \{F_1,\dots,F_N\}$ and restricting to those $\gamma \in \mathbf{B}(\mu,\nu)$ whose set of focal elements satisfies

$$\mathcal{F}_\gamma \subseteq \mathcal{F}_\mu \otimes \mathcal{F}_\nu := \{E \times F \,:\, E \in \mathcal{F}_\mu, F \in \mathcal{F}_\nu\}. \tag{6}$$

This allows us to introduce the following set of joint belief functions with given marginals

$$\overline{\mathbf{B}}(\mu,\nu) = \{\gamma \in \mathbf{B}(\mu,\nu) \,:\, \mathcal{F}_\gamma \subseteq \mathcal{F}_\mu \otimes \mathcal{F}_\nu\}. \tag{7}$$

The following proposition lists a set of properties of $\overline{\mathbf{B}}(\mu,\nu)$ that are straightforward to demonstrate and generalizes a well-known result in probability.

Proposition 1. *The following statements hold:*

(i) $\overline{\mathbf{B}}(\mu,\nu) \subseteq \mathbf{B}(\mu,\nu)$*;*
(ii) $\overline{\mathbf{B}}(\mu,\nu)$ *is a non-empty closed convex subset of* $[0,1]^{2^{\mathcal{X}\times\mathcal{Y}}}$*, endowed with the product topology.*

In general, it holds that $\overline{\mathbf{B}}(\mu,\nu) \subset \mathbf{B}(\mu,\nu)$, since we can find a joint belief function $\gamma \in \mathbf{B}(\mu,\nu)$ that is not in $\overline{\mathbf{B}}(\mu,\nu)$, as the following example shows.

Example 1. Consider $\mathcal{X} = \mathcal{Y} = \{a, b\}$, $\mu = \delta_{\mathcal{X}}$, $\nu = \delta_{\mathcal{Y}}$, and denote $xy := (x, y)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$. Taking $\gamma = \delta_{\{aa,ba,bb\}}$ we have that $\gamma_{\widetilde{|2^{\mathcal{X}}}} = \mu$ and $\gamma_{\widetilde{|2^{\mathcal{Y}}}} = \nu$, since

- $\gamma(\emptyset \times \mathcal{Y}) = \gamma(\mathcal{X} \times \emptyset) = 0$;
- $\gamma(\{a\} \times \mathcal{Y}) = \gamma(\{aa, ab\}) = \gamma(\{aa, ba\}) = \gamma(\mathcal{X} \times \{a\}) = 0$;
- $\gamma(\{b\} \times \mathcal{Y}) = \gamma(\{ba, bb\}) = \gamma(\{ab, bb\}) = \gamma(\mathcal{X} \times \{b\}) = 0$;
- $\gamma(\mathcal{X} \times \mathcal{Y}) = 1$.

Then, $\mathcal{F}_\gamma = \{\{aa, ba, bb\}\}$ is not contained in $\mathcal{F}_\mu \otimes \mathcal{F}_\nu = \{\mathcal{X} \times \mathcal{Y}\}$, therefore $\gamma \in \mathbf{B}(\mu, \nu)$ but $\gamma \notin \overline{\mathbf{B}}(\mu, \nu)$. ♦

The main advantage of $\overline{\mathbf{B}}(\mu, \nu)$ is that its elements can be characterized through copula theory (see [8, 11]). We recall that a *bivariate copula* (see, e.g., [12]) is a function $\mathsf{C} : [0, 1]^2 \to [0, 1]$ satisfying the following properties: for all $x, y, x', y' \in [0, 1]$ with $x \leq x'$ and $y \leq y'$, it holds that

(i) $\mathsf{C}(x, 0) = \mathsf{C}(0, y) = 0$;
(ii) $\mathsf{C}(x, 1) = x$ and $\mathsf{C}(1, y) = y$;
(iii) $\mathsf{C}(x', y') + \mathsf{C}(x, y) - \mathsf{C}(x, y') - \mathsf{C}(x', y) \geq 0$.

As is well-known, all copulas are bounded by the Łukasiewicz (C_L) and the minimum (C_M) copulas, i.e., every copula C satisfies, for all $x, y \in [0, 1]$,

$$\max(0, x + y - 1) =: \mathsf{C}_L(x, y) \leq \mathsf{C}(x, y) \leq \mathsf{C}_M(x, y) := \min(x, y).$$

In order to generate an element of $\overline{\mathbf{B}}(\mu, \nu)$ through a copula C (see [8, 11]), referring to $\mathcal{F}_\mu$ and $\mathcal{F}_\nu$ we have to choose a permutation σ of $\{1, \ldots, M\}$ and a permutation τ of $\{1, \ldots, N\}$, and consider the corresponding order relations $E_{\sigma(1)} \leq_\sigma \cdots \leq_\sigma E_{\sigma(M)}$ and $F_{\tau(1)} \leq_\tau \cdots \leq_\tau F_{\tau(N)}$.

Given μ, ν, σ, τ, and C, we can define a function $m^{\mu,\nu,\sigma,\tau,\mathsf{C}} : 2^{\mathcal{X} \times \mathcal{Y}} \to [0, 1]$, by setting $m^{\mu,\nu,\sigma,\tau,\mathsf{C}}(A) = 0$ for all $A \in 2^{\mathcal{X} \times \mathcal{Y}} \setminus (\mathcal{F}_\mu \otimes \mathcal{F}_\nu)$, while for all $E_{\sigma(i)} \times F_{\tau(j)} \in \mathcal{F}_\mu \otimes \mathcal{F}_\nu$ we set

$$\begin{aligned} m^{\mu,\nu,\sigma,\tau,\mathsf{C}}(E_{\sigma(i)} \times F_{\tau(j)}) = {} & \mathsf{C}\left(M^\sigma_\mu(i), M^\tau_\nu(j)\right) + \mathsf{C}\left(M^\sigma_\mu(i-1), M^\tau_\nu(j-1)\right) \\ & - \mathsf{C}\left(M^\sigma_\mu(i), M^\tau_\nu(j-1)\right) - \mathsf{C}\left(M^\sigma_\mu(i-1), M^\tau_\nu(j)\right), \end{aligned}$$

where $M^\sigma_\mu(i) = \sum_{k \leq i} m_\mu(E_{\sigma(k)})$, $M^\tau_\nu(j) = \sum_{k \leq j} m_\nu(F_{\tau(k)})$, and summations over an empty index set are assumed to be 0. It is readily proven that $m^{\mu,\nu,\sigma,\tau,\mathsf{C}}$ is the Möbius inverse of a belief function belonging to $\overline{\mathbf{B}}(\mu, \nu)$. By an adaptation of the classical Sklar's theorem [12], we get the following result.

Theorem 1. *Let σ and τ be permutations of $\{1, \ldots, M\}$ and $\{1, \ldots, N\}$, respectively. It holds that $\gamma \in \overline{\mathbf{B}}(\mu, \nu)$ if and only if there exists a bivariate copula C such that $m_\gamma = m^{\mu,\nu,\sigma,\tau,\mathsf{C}}$.*

Since $\overline{\mathbf{B}}(\mu, \nu)$ is a closed and convex set, it is still determined by its set of extreme points $\mathrm{ext}(\overline{\mathbf{B}}(\mu, \nu))$, the latter being finite. Following [15], we need the following definition to characterize $\mathrm{ext}(\overline{\mathbf{B}}(\mu, \nu))$.

Definition 1. *Let $\gamma \in \overline{\mathbf{B}}(\mu,\nu)$ and let $\mathbf{M}_\gamma = (m_{\gamma_{ij}})_{\substack{i=1,\ldots,N\\ j=1,\ldots,M}}$ with $m_{\gamma_{ij}} := m_\gamma(E_i \times F_j)$, for all $E_i \times F_j \in \mathcal{F}_\mu \otimes \mathcal{F}_\nu$. An ordered sequence $[m_{\gamma_{i_1 j_1}}, \ldots, m_{\gamma_{i_k j_k}}]$ in $\mathbf{M}_\gamma$ is said to be a* **loop** *in $\mathbf{M}_\gamma$ if*

- *k is even;*
- *$i_1 = i_2,\ i_3 = i_4,\ \cdots,\ i_{k-1} = i_k$;*
- *$j_2 = j_3,\ j_4 = j_5,\ \cdots,\ j_k = j_1$;*
- *the pairs (i_r, j_r) for $r = 1, \ldots, k$ are all distinct.*

By a straightforward adaptation of the proof of Theorem 2.9 in [15], the following theorem is easily established.

Theorem 2 (Characterization of $\mathrm{ext}(\overline{\mathbf{B}}(\mu,\nu))$). *Let $\gamma \in \overline{\mathbf{B}}(\mu,\nu)$ and let $\mathbf{M}_\gamma$ as in Definition 1. The following statements are equivalent:*

(i) γ is not an extreme point of $\overline{\mathbf{B}}(\mu,\nu)$;
(ii) there is a positive loop in $\mathbf{M}_\gamma$, i.e., every member of the loop is positive.
(iii) there exists a submatrix $\overline{\mathbf{M}}_\gamma$ of $\mathbf{M}_\gamma$ having the property that every row and column of $\overline{\mathbf{M}}_\gamma$ has at least two positive elements;
(iv) there exists a square submatrix $\widetilde{\mathbf{M}}_\gamma$ of $\mathbf{M}_\gamma$ having the property that every row and column of $\widetilde{\mathbf{M}}_\gamma$ contains at least two positive entries;
(v) there exists a square submatrix $\mathbf{M}^*_\gamma$ of $\mathbf{M}_\gamma$ of order $k \times k$ for some $k \geq 1$ having the property that the number of positive elements in $\mathbf{M}^*_\gamma$ is at least $2k$.

By Theorem 1, we known that, for fixed permutations σ and τ, elements of $\overline{\mathbf{B}}(\mu,\nu)$ can be obtained varying the copula C. We could think that $\mathrm{ext}(\overline{\mathbf{B}}(\mu,\nu))$ can be generated by varying the permutations σ and τ, and limiting to the two extreme copulas C_L and C_M. The following example shows that, in general, some extreme points could not be obtained from C_L and C_M.

Example 2. Let $\mathcal{X} = \{x_1, x_2, x_3, x_4\}$, $\mathcal{Y} = \{y_1, y_2, y_3\}$, and consider the marginal belief functions μ, ν with sets of focal elements $\mathcal{F}_\mu = \{E_1 = \{x_1, x_2\}, E_2 = \{x_1, x_3\}, E_3 = \{x_1, x_4\}, E_4 = \{x_3, x_4\}\}$, $\mathcal{F}_\nu = \{F_1 = \{y_1, y_2\}, F_2 = \{y_2, y_3\}, F_3 = \mathcal{Y}\}\}$, and Möbius inverses

$\mathcal{F}_\mu$	E_1	E_2	E_3	E_4
m_μ	$\frac{10}{20}$	$\frac{2}{20}$	$\frac{4}{20}$	$\frac{4}{20}$

$\mathcal{F}_\nu$	F_1	F_2	F_3
m_ν	$\frac{8}{20}$	$\frac{4}{20}$	$\frac{8}{20}$

Denote by Σ the set of permutations of $\{1, \ldots, 4\}$, by T the set of permutations of $\{1, 2, 3\}$, and define the set of joint belief functions generated by the Łukasiewicz or the minimum copulas

$$\Gamma = \{\gamma \in \overline{\mathbf{B}}(\mu,\nu) \,:\, m_\gamma = m^{\mu,\nu,\sigma,\tau,\mathsf{C}_L} \text{ or } m_\gamma = m^{\mu,\nu,\sigma,\tau,\mathsf{C}_M},\ \sigma \in \Sigma,\ \tau \in T\}.$$

It turns out that $|\mathrm{ext}(\overline{\mathbf{B}}(\mu,\nu))| = 34$, $|\Gamma| = 32$, and $\Gamma \subset \mathrm{ext}(\overline{\mathbf{B}}(\mu,\nu))$, with $\mathrm{ext}(\overline{\mathbf{B}}(\mu,\nu)) \setminus \Gamma = \{\gamma_1, \gamma_2\}$, where the matrices corresponding to γ_1, γ_2 are

$$\mathbf{M}_{\gamma_1} = \begin{pmatrix} \frac{4}{20} & \frac{2}{20} & \frac{4}{20} \\ 0 & \frac{2}{20} & 0 \\ \frac{4}{20} & 0 & 0 \\ 0 & 0 & \frac{4}{20} \end{pmatrix} \quad \text{and} \quad \mathbf{M}_{\gamma_2} = \begin{pmatrix} \frac{4}{20} & \frac{2}{20} & \frac{4}{20} \\ 0 & \frac{2}{20} & 0 \\ 0 & 0 & \frac{4}{20} \\ \frac{4}{20} & 0 & 0 \end{pmatrix}.$$

A direct verification shows that there are no permutations in Σ and T, such that m_{γ_1} and m_{γ_2} can be obtained through C_L or C_M. ♦

The restriction to sets of joint belief functions of the form (7) allows us to prove the following version of the famous gluing lemma [20].

Lemma 1 (Gluing Lemma). *Let $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ be finite sets and let μ, ν and η be belief functions on $2^{\mathcal{X}}$, $2^{\mathcal{Y}}$ and $2^{\mathcal{Z}}$, respectively. If $\gamma_1 \in \overline{\mathbf{B}}(\mu,\nu)$ and $\gamma_2 \in \overline{\mathbf{B}}(\nu,\eta)$, then there exist a joint belief function on $(\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}, 2^{\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}})$ with marginals γ_1 and γ_2, on $\widetilde{2^{\mathcal{X}\times\mathcal{Y}}}$ and $\widetilde{2^{\mathcal{Y}\times\mathcal{Z}}}$, and focal elements in*

$$\mathcal{F}_\mu \otimes \mathcal{F}_\nu \otimes \mathcal{F}_\eta := \{E \times F \times G \,:\, E \in \mathcal{F}_\mu, F \in \mathcal{F}_\nu, G \in \mathcal{F}_\eta\}.$$

4 Optimal Transport in Dempster-Shafer Theory

Let μ and ν be two probability measures on the finite spaces $\mathcal{X} = \{x_1,\dots,x_m\}$ and $\mathcal{Y} = \{y_1,\dots,y_n\}$, respectively. For a cost function $c : \mathcal{X}\times\mathcal{Y} \to [0,+\infty)$, the classical Kantorovich *optimal transport problem* between μ and ν with respect to the cost c is given by

$$\mathrm{OT}(\mu,\nu,c) = \min_{\pi\in\mathbf{P}(\mu,\nu)} \int c(x,y)\mathrm{d}\pi(x,y), \tag{8}$$

where $\mathbf{P}(\mu,\nu)$ is the set of joint probability measures on $\mathcal{X}\times\mathcal{Y}$ with first marginal equal to μ and second marginal equal to ν. Assuming that $\mathcal{X} = \mathcal{Y}$ and fixing $c : \mathcal{X}^2 \to [0,+\infty)$ which is a metric on $\mathcal{X}$, then

$$d_{\mathcal{W}}(\mu,\nu) = \mathrm{OT}(\mu,\nu,c) \tag{9}$$

defines the so-called *Wasserstein distance (of order* 1*)* (see, e.g., [20]) on the set of probability measures on $\mathcal{X}$.

Our aim is to generalize (8) and (9) in the context of belief functions. Referring to the set of joint belief functions (7), we can introduce the following pessimistic and optimistic versions of optimal transport in Dempster-Shafer theory.

Definition 2. *Let μ and ν be two belief functions on $2^{\mathcal{X}}$ and $2^{\mathcal{Y}}$, respectively, with $\mathcal{X} = \{x_1,\dots,x_m\}$ and $\mathcal{Y} = \{y_1,\dots,y_n\}$. For a cost function $c : \mathcal{X}\times\mathcal{Y} \to [0,+\infty)$, we define*

– *the* **pessimistic Dempster-Shafer optimal transport problem** *between μ and ν with respect to the cost c as*

$$\begin{aligned}\mathrm{DSOT}(\mu,\nu,c) &= \min_{\gamma\in\overline{\mathbf{B}}(\mu,\nu)} \oint c(x,y)\mathrm{d}\gamma(x,y)\\ &= \min_{\gamma\in\overline{\mathbf{B}}(\mu,\nu)} \sum_{E\times F\in\mathcal{F}_\gamma} C(E\times F)m_\gamma(E\times F),\end{aligned} \tag{10}$$

where $C : \mathcal{F}_\gamma \to [0,+\infty)$ is such that $C(E\times F) = \min_{(x,y)\in E\times F} c(x,y)$;

– *the* **optimistic Dempster-Shafer optimal transport problem** *between μ and ν with respect to the cost c as*

$$\begin{aligned}\overline{\mathrm{DSOT}}(\mu,\nu,c) &= \min_{\gamma\in\overline{\mathbf{B}}(\mu,\nu)} \oint c(x,y)\mathrm{d}\overline{\gamma}(x,y)\\ &= \min_{\gamma\in\overline{\mathbf{B}}(\mu,\nu)} \sum_{E\times F\in\mathcal{F}_\gamma} \overline{C}(E\times F)m_\gamma(E\times F),\end{aligned} \tag{11}$$

where $\overline{\gamma}$ is the dual plausibility function of $\gamma\in\overline{\mathbf{B}}(\mu,\nu)$ and $\overline{C}:\mathcal{F}_\gamma\to[0,+\infty)$ is such that $\overline{C}(E\times F) = \max_{(x,y)\in E\times F} c(x,y)$.

Recalling (1) and (2), we get that $\mathrm{DSOT}(\mu,\nu,c)$ and $\overline{\mathrm{DSOT}}(\mu,\nu,c)$ implement a *minimin* and a *minimax* decision rule, respectively.

The following toy example shows a possible situation where $\mathrm{DSOT}(\mu,\nu,c)$ and $\overline{\mathrm{DSOT}}(\mu,\nu,c)$ could be relevant.

Example 3. Consider a population of people that pay taxes and use public services, where individuals are categorized in retired (R) and non-retired (N). The variable X = "tax payer" has known distribution on retired and non-retired people, while for Y = "public service user" it is only known that retired people using public services are at least 25%. The cost function is defined respecting a sort of social equity and prizes categories paying for public service they use.

Referring to X and Y ranging in $\mathcal{X}=\mathcal{Y}=\{R,N\}$, the previous situation can be described by the following marginal belief functions and cost function:

– $\mathcal{F}_\mu = \{\{R\},\{N\}\}$ and $m_\mu(\{R\}) = m_\mu(\{N\}) = \frac{1}{2}$;
– $\mathcal{F}_\nu = \{\{R\},\mathcal{Y}\}$ and $m_\nu(\{R\}) = \frac{1}{4}$ and $m_\nu(\mathcal{Y}) = \frac{3}{4}$;
– $c(x,y) = \begin{cases} €1000, & \text{if } x\neq y,\\ €500, & \text{otherwise.}\end{cases}$

Given the previous information, we get

$$\mathrm{DSOT}(\mu,\nu,c) = \min_{\gamma\in\overline{\mathbf{B}}(\mu,\nu)} \oint c(x,y)\mathrm{d}\gamma(x,y) = €500,$$
$$\overline{\mathrm{DSOT}}(\mu,\nu,c) = \min_{\gamma\in\overline{\mathbf{B}}(\mu,\nu)} \oint c(x,y)\mathrm{d}\overline{\gamma}(x,y) = €1000,$$

that is what we expect to earn from a person, from the government's point of view, referring to a pessimistic and an optimistic approach, respectively. ♦

5 Choquet-Wasserstein Pseudo-Distances

Assume $\mathcal{X}=\mathcal{Y}=\{x_1,\dots,x_m\}$ and $c:\mathcal{X}^2\to[0,+\infty)$ is a metric on $\mathcal{X}$: a typical choice for c is the *discrete metric* defined, for all $x,y\in\mathcal{X}$, as

$$c_d(x,y) = \begin{cases} 1, \text{ if } x\neq y,\\ 0, \text{ if } x = y.\end{cases}$$

Another common choice is to take the *absolute value metric* $c_a(x,y) = |x-y|$, provided $\mathcal{X} \subseteq \mathbb{R}$, or the *Euclidean metric* $c_e(x,y) = \|x-y\|_2$, provided $\mathcal{X} \subseteq \mathbb{R}^d$.

The optimal transport problems in Dempster-Shafer theory presented in Definition 2 give naturally rise to two Choquet-Wasserstein pseudo-distances. We first analyze the pessimistic pseudo-distance.

Definition 3. *Let c be a metric on $\mathcal{X}$. Given two belief functions μ and ν on $2^{\mathcal{X}}$, we define their* **pessimistic Choquet-Wasserstein pseudo-distance** *as*

$$d_{\mathcal{CW}}(\mu,\nu) = \mathrm{DSOT}(\mu,\nu,c) = \min_{\gamma \in \overline{\mathbf{B}}(\mu,\nu)} \oint c(x,y)\mathrm{d}\gamma(x,y).$$

We point out that $d_{\mathcal{CW}}$ is a non-negative real-valued function, defined on the set of pairs of belief functions on $2^{\mathcal{X}}$. Nevertheless, $d_{\mathcal{CW}}$ is not a metric on the whole set of belief function on $2^{\mathcal{X}}$, as it may fail the positivity property and the triangular inequality.

Example 4 (Positivity). Let $\mathcal{X} = \{a, b\}$ and take the discrete metric $c = c_d$. For $\mu = \delta_{\mathcal{X}}$ and $\nu = \delta_{\{a\}}$, we have that $\mathcal{F}_\mu = \{\mathcal{X}\}$, $\mathcal{F}_\nu = \{\{a\}\}$ and $\mathcal{F}_\mu \otimes \mathcal{F}_\nu = \{\mathcal{X} \times \{a\}\}$. Then, since $C(\mathcal{X} \times \{a\}) = 0$, we get

$$d_{\mathcal{CW}}(\mu,\nu) = \min_{\gamma \in \overline{\mathbf{B}}(\mu,\nu)} C(\mathcal{X} \times \{a\}) m_\gamma(\mathcal{X} \times \{a\}) = 0,$$

that is $\mu \neq \nu$ does not imply $d_{\mathcal{CW}}(\mu,\nu) > 0$. ♦

Example 5 (Triangular Inequality). Let $\mathcal{X} = \{a, b\}$ and take the discrete metric $c = c_d$. For $\mu = \delta_{\{a\}}$, $\nu = \delta_{\{b\}}$ and $\eta = \delta_{\mathcal{X}}$, we have that $\mathcal{F}_\mu \otimes \mathcal{F}_\nu = \{\{a\} \times \{b\}\}$, $\mathcal{F}_\mu \otimes \mathcal{F}_\eta = \{\{a\} \times \mathcal{X}\}$ and $\mathcal{F}_\eta \otimes \mathcal{F}_\nu = \{\mathcal{X} \times \{b\}\}$. Since $C(\{a\} \times \{b\}) = 1$ and $C(\{a\} \times \mathcal{X}) = C(\mathcal{X} \times \{b\}) = 0$, we get that

$$\begin{aligned} d_{\mathcal{CW}}(\mu,\nu) &= \min_{\gamma_1 \in \overline{\mathbf{B}}(\mu,\nu)} C(\{a\} \times \{b\}) m_{\gamma_1}(\{a\} \times \{b\}) = 1, \\ d_{\mathcal{CW}}(\mu,\eta) &= \min_{\gamma_2 \in \overline{\mathbf{B}}(\mu,\eta)} C(\{a\} \times \mathcal{X}) m_{\gamma_2}(\{a\} \times \mathcal{X}) = 0, \\ d_{\mathcal{CW}}(\eta,\nu) &= \min_{\gamma_3 \in \overline{\mathbf{B}}(\eta,\nu)} C(\mathcal{X} \times \{b\}) m_{\gamma_3}(\mathcal{X} \times \{b\}) = 0, \end{aligned}$$

and so $d_{\mathcal{CW}}(\mu,\nu) > d_{\mathcal{CW}}(\mu,\eta) + d_{\mathcal{CW}}(\eta,\nu)$. ♦

The following theorem lists the properties satisfied by $d_{\mathcal{CW}}$, which show that $d_{\mathcal{CW}}$ is a *distance* or *dissimilarity* function on the set of belief functions on $2^{\mathcal{X}}$, according to the terminology of [5].

Theorem 3. *Let c be a metric on $\mathcal{X}$. For all belief functions μ, ν on $2^{\mathcal{X}}$, the function $d_{\mathcal{CW}}$ satisfies:*

(i) $d_{\mathcal{CW}}(\mu,\nu) \geq 0$;
(ii) $\mu = \nu \implies d_{\mathcal{CW}}(\mu,\nu) = 0$;
(iii) $d_{\mathcal{CW}}(\mu,\nu) = d_{\mathcal{CW}}(\nu,\mu)$.

Analogously, we can define an optimistic pseudo-distance as follows.

Definition 4. *Let c be a metric on $\mathcal{X}$. Given two belief functions μ and ν on $2^{\mathcal{X}}$, we define their* **optimistic Choquet-Wasserstein pseudo-distance** *as*

$$\overline{d}_{\mathcal{CW}}(\mu,\nu) = \overline{\mathrm{DSOT}}(\mu,\nu,c) = \min_{\gamma\in\overline{\mathbf{B}}(\mu,\nu)} \oint c(x,y)\mathrm{d}\overline{\gamma}(x,y),$$

where $\overline{\gamma}$ is the dual plausibility function of γ.

As before, $\overline{d}_{\mathcal{CW}}$ is not a metric, as it may fail the reflexivity property.

Example 6 (Reflexivity). Let $\mathcal{X} = \{a, b\}$ and take the discrete metric $c = c_d$. Let $\mu = \nu = \delta_{\mathcal{X}}$, for which $\mathcal{F}_\mu = \mathcal{F}_\nu = \{\mathcal{X}\}$ and $\mathcal{F}_\mu \otimes \mathcal{F}_\nu = \{\mathcal{X}\times\mathcal{X}\}$. Since $\overline{C}(\mathcal{X}\times\mathcal{X}) = 1$, we get that

$$\overline{d}_{\mathcal{CW}}(\mu,\nu) = \min_{\gamma\in\overline{\mathbf{B}}(\mu,\nu)} \overline{C}(\mathcal{X}\times\mathcal{X})m_\gamma(\mathcal{X}\times\mathcal{X}) = 1,$$

thus $\mu=\nu$ does not imply $\overline{d}_{\mathcal{CW}}(\mu,\nu) = 0$. ♦

The following theorem lists the properties satisfied by $\overline{d}_{\mathcal{CW}}$, which show that $\overline{d}_{\mathcal{CW}}$ is a *metric-like* function according to the terminology of [5].

Theorem 4. *Let c be a metric on $\mathcal{X}$. For all belief functions μ,ν,η on $2^{\mathcal{X}}$, the function $\overline{d}_{\mathcal{CW}}$ satisfies:*

(i) $\overline{d}_{\mathcal{CW}}(\mu,\nu) \geq 0$;
(ii) $\overline{d}_{\mathcal{CW}}(\mu,\nu) = \overline{d}_{\mathcal{CW}}(\nu,\mu)$;
(iii) $\overline{d}_{\mathcal{CW}}(\mu,\nu) = 0 \implies \mu=\nu$;
(iv) $\overline{d}_{\mathcal{CW}}(\mu,\nu) \leq \overline{d}_{\mathcal{CW}}(\mu,\eta) + \overline{d}_{\mathcal{CW}}(\eta,\nu)$.

Given an arbitrary belief function μ, both pseudo-distances $d_{\mathcal{CW}}$ and $\overline{d}_{\mathcal{CW}}$, can be used to find a "closest" belief function ν belonging to a distinguished subclass of belief functions. A very popular choice in applications is given by the class of *at most k-additive* belief functions [9, 10], obtained as the union of h-additive ones, for $h = 1,\ldots,k$, that we denote by $\mathbf{A}_k(\mathcal{X})$. The set $\mathbf{A}_k(\mathcal{X})$ is easily seen to be a closed and convex set of belief functions on $2^{\mathcal{X}}$. We notice that an element ν of $\mathbf{A}_k(\mathcal{X})$ has at most $\sum_{h=1}^{k}\binom{m}{h}$ focal elements thus we can refer to $\mathcal{F}_\nu = \{F : |F| \leq k\}$, as the maximal set of focal elements. For simplicity, here we assume $c = c_d$.

To find a $d_{\mathcal{CW}}$-minimal element of $\mathbf{A}_k(\mathcal{X})$ with respect to μ, we need to solve the problem

$$\nu^* \in \operatorname*{arg\,min}_{\nu\in\mathbf{A}_k(\mathcal{X})} d_{\mathcal{CW}}(\mu,\nu) = \operatorname*{arg\,min}_{\nu\in\mathbf{A}_k(\mathcal{X})} \mathrm{DSOT}(\mu,\nu,c_d), \tag{12}$$

which is generally not easy to attack as it involves a double minimization. Following [2, 13], to solve (12) we consider the *(negative) entropy*

$$H(\gamma) = \sum_{E_i\times F_j\in\mathcal{F}_\mu\otimes\mathcal{F}_\nu} m_\gamma(E_i\times F_j)(\ln(m_\gamma(E_i\times F_j)) - 1), \tag{13}$$

and define, for $\lambda > 0$, the *entropic regularization*

$$\mathrm{DSOT}_\lambda(\mu, \nu, c_d) = \min_{\gamma \in \overline{\mathbf{B}}(\mu,\nu)} \left\{ \oint c_d(x, y) \mathrm{d}\gamma(x, y) + \lambda H(\gamma) \right\}. \tag{14}$$

It turns out that, for a fixed belief function μ on $2^{\mathcal{X}}$, $\mathrm{DSOT}_\lambda(\mu, \nu, c_d)$ is a strictly convex function of ν in $\mathbf{A}_k(\mathcal{X})$ admitting a unique optimizer

$$\nu_\lambda^* = \operatorname*{arg\,min}_{\nu \in \mathbf{A}_k(\mathcal{X})} \mathrm{DSOT}_\lambda(\mu, \nu, c_d), \tag{15}$$

moreover, for $\lambda \to 0$, it can be shown that ν_λ^* converges pointwise to an optimizer ν^* of the original problem (12). Thus, for a sufficiently small value of λ, we get a good approximation of the searched optimizer.

Analogously, we can reformulate problem (12) and its entropic regularization (14) with respect to $\overline{d}_{\mathcal{CW}}$ and find

$$\nu_\lambda^{**} = \operatorname*{arg\,min}_{\nu \in \mathbf{A}_k(\mathcal{X})} \overline{\mathrm{DSOT}}_\lambda(\mu, \nu, c_d). \tag{16}$$

Both problems (15) and (16) and can be faced by adapting *Dykstra's algorithm* [2, 13].

Example 7. Let $\mathcal{X} = \{x_1, x_2, x_3\}$ and consider the belief function μ on $2^{\mathcal{X}}$ with set of focal elements $\mathcal{F}_\mu = \{E_1 = \{x_1\}, E_2 = \{x_1, x_3\}, E_3 = \mathcal{X}\}$ and Möbius inverse $m_\mu(E_1) = \frac{3}{6}$, $m_\mu(E_2) = \frac{1}{6}$, and $m_\mu(E_3) = \frac{2}{6}$. Table 1 shows the Möbius

$\mathcal{F}_\nu$	$\{x_1\}$	$\{x_2\}$	$\{x_3\}$	$\{x_1, x_2\}$	$\{x_1, x_3\}$	$\{x_2, x_3\}$
$m_{\nu^*_{0.05}}$	0.255556	0.055556	0.088889	0.255556	0.255556	0.088889
$m_{\nu^{**}_{0.05}}$	0.583333	0.083333	0.083333	0.083333	0.083333	0.083333

Table 1. Möbius inverses of ν_λ^* and ν_λ^{**} in (15) and (16) for $k = 2$ and $\lambda = 0.05$.

inverses of the optimal solutions of (15) and (16) for $k = 2$ and $\lambda = 0.05$, computed by adapting Dykstra's algorithm. ♦

6 Conclusions

We considered the marginal problem in Dempster-Shafer theory and provided an analysis of a suitable subset of joint belief functions with given marginals, introduced in [8, 17]. Restricting to such a set of joint belief functions, we formulated a pessimistic and an optimistic Kantorovich-like optimal transport problem and defined a pessimistic and an optimistic Choquet-Wasserstein pseudo-distance. Finally, we introduced an entropic regularization to compute a "closest" at most k-additive belief function of a given belief function, according to the two pseudo-distances. In the same spirit, one can address probability-possibility transformations for metrology (see, e.g., [7]). The aim of future research is to further analyze the derived Choquet-Wasserstein pseudo-distances, focusing on their application to machine learning algorithms like WGANs [1], so as to convey ambiguity.

Acknowledgments. We acknowledge the support of the PRIN 2022 project “Models for dynamic reasoning under partial knowledge to make interpretable decisions” (Project number: `2022AP3B3B`, CUP Master: `J53D23004340006`, CUP: `B53D23009860006`) funded by the European Union – Next Generation EU.

References

1. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: Proc. 34th International Conference on Machine Learning. PMLR, vol. 70, pp. 214–223 (2017)
2. Blanchet, A., Carlier, G., Nenna, L.: Computation of Cournot-Nash equilibria by entropic regularization. Vietnam J. Math. **46**, 15–31 (2017)
3. Bronevich, A., Rozenberg, I.: The measurement of relations on belief functions based on the Kantorovich problem and the Wasserstein metric. Int. J. Approx. Reas. **131**, 108–135 (2021)
4. Dempster, A.: Upper and Lower Probabilities Induced by a Multivalued Mapping. Ann. Math. Stat. **38**(2), 325–339 (1967)
5. Deza, M., Deza, E.: Encyclopedia of Distances. Springer, 4th edn. (2016)
6. Ellsberg, D.: Risk, Ambiguity, and the Savage Axioms. Quart. J. Econom. **75**(4), 643–669 (1961)
7. Ferrero, A., Prioli, M., Salicone, S., Vantaggi, B.: A 2-D Metrology-Sound Probability–Possibility Transformation. IEEE Trans. Instr. Meas. **62**(5), 982–990 (2013)
8. Ferson, S., Nelsen, R., Hajagos, J., Berleant, D., Zhang, J., Tuckerand, W., Ginzburg, L., Oberkampf, W.: Dependence in probabilistic modeling, dempster-shafer theory, and probability bounds analysis. Tech. rep., Sandia Report (2004)
9. Grabisch, M.: k-order additive discrete fuzzy measures and their representation. Fuzzy Sets Sys. **92**(2), 167–189 (1997)
10. Grabisch, M.: Set Functions, Games and Capacities in Decision Making. Theory and Decision Library C, Springer (2016)
11. Malinowski, R., Destercke, S.: Copulas, Lower Probabilities and Random Sets: How and When to Apply Them? In: García-Escudero, L., Gordaliza, A., Mayo, A., Lubiano, M., Gil, M., Grzegorzewski, P., Hryniewicz, O. (eds.) Building Bridges between Soft and Statistical Methodologies for Data Science. pp. 271–278. Springer International Publishing, Cham (2023)
12. Nelsen, R.: An Introduction to Copulas. Springer, 2nd edn. (2006)
13. Peyré, G.: Entropic Approximation of Wasserstein Gradient Flows. SIAM J. Imag. Sci. **8**(4), 2323–2351 (2015)
14. Peyré, G., Cuturi, M.: Computational optimal transport. Found. Trends Machine Learn. **11**(5-6), 355–607 (2019)
15. Rao, M.B., Subramanyam, K.: The structure of some classes of bivariate distributions and some applications. Comp. Stat. Data Anal. **10**(2), 175–187 (1990)
16. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press, Princeton, NJ (1976)
17. Studený, M.: Marginal problem in different calculi of AI. In: Bouchon-Meunier, B., Yager, R., Zadeh, L. (eds.) Advances in Intelligent Computing — IPMU ’94. pp. 348–359. Springer Berlin Heidelberg, Berlin, Heidelberg (1995)
18. Torra, V.: The transport problem for non-additive measures. Eur. J. Oper. Res. **311**(2), 679–689 (2023)
19. Villani, C.: Topics in Optimal Transport. American Mathematical Society (2003)
20. Villani, C.: Optimal Transport: Old and New. Springer-Verlag, Berlin, Heidelberg (2009)