Interpreting evidential distances by connecting them to partial orders: Application to belief function approximation

John Klein, Sebastien Destercke, Olivier Colot

TOWARDS A SEMANTIC INTERPRETATION OF EVIDENTIAL DISTANCES: THE CASE OF BELIEF FUNCTIONS APPROXIMATION

JOHN KLEIN, SEBASTIEN DESTERCKE AND OLIVIER COLOT

ABSTRACT. The many distances defined in evidence theory provide instrumental tools to analyze and compare mass functions: they have been proposed to measure conflict, dependence or similarity in different fields (information fusion, risk analysis, machine learning). Many of their mathematical properties have been studied in the past years, yet a remaining question is to know what distance to choose in a particular problem. As a step towards answering this question, we propose to interpret distances by looking at their consistency with partial orders possessing a clear semantic. We focus on the case of informational partial order and on the problem of approximating initial belief functions by simpler ones. Doing so, we study which distances can be used to measure the difference of informational content between two mass functions, and which distances cannot.

Keywords: Belief functions, distance, metric, partial order, semantic, convex optimization

1. INTRODUCTION

The theory of belief functions is a flexible framework to model uncertainty in the presence of imprecision. This framework mixes set and probabilistic representations. It was initially proposed to model imprecise statistical observations [1], and this initial work was then extended [2] to include subjective and non-statistical uncertainty (*e.g.*, when a variable has a fixed, yet ill-known value). This latter view was then pursued by Smets [3], who dissociated belief functions from any probabilistic interpretation. They include many other representations proposed in the literature, such as sets, probability measures or possibility measures. In this paper, we will use the term evidence theory as a generic term for frameworks relying on belief functions.

Among the tools developed to work with belief functions, distances have recently received a growing attention. They have been proposed as tools to achieve various tasks: measuring conflict [4, 5], measuring dependencies [6], learning models from data [7, 8], or belief function approximation [9, 10, 11, 12, 13, 14, 15]. Jousselme and Maupin [16] surveyed evidential distances in order to classify them with respect to their mathematical properties and to show some correlated behaviors among them. Following Jousselme and Maupin's analysis, Loudahi *et al.*[17, 18] formalized some properties with intuitive interpretations in the framework of evidence theory. Despite these efforts, providing evidential distances with clear interpretations remains an open problem.

In this paper, we start by proposing a new answer to this problem: we interpret a distance by its compatibility or incompatibility with some partial order possessing a clear semantic. We then focus on the case of belief function approximations, in which partial orders related to informative content play a specific role. The combination of both distances and orders is very interesting in this problem, as the partial orders allow us to select those distances consistent with the approximation

problem, while the use of specific distances (within the selected subset) allows us to take advantage of their mathematical properties to find unique solutions (when partial orders only offers sets of incomparable solutions). Finally, we end with some discussion about the presented results and ideas.

In Section 2, we will briefly review the basic concepts of the theory of belief functions. Section 3 will recall the notions of distances and partial orders, as well as describe our main idea to interpret distances according to their compatibility with partial orders. Section 4 will then focus on the problem of approximating belief functions, using partial orders of informative content to select and study adequate distances. In particular, we will show that only some of the classical distances are consistent with those partial orders, and that the choice of distance can be further refined if we impose the approximation problem to have a unique solution. Section 5 discusses some related points and possible extensions.

2. Basics of evidence theory

This section contains a reminder of evidence theory notions used in this paper. More details about the various tools used in evidence theory can be found in [19], for example. After providing this necessary background, we give a more detailed presentation of partial orders and distances used in evidence theory.

Let $\Omega = \{\omega_1, \ldots, \omega_n\}$ be a finite space over which a given ill-known variable θ takes its values. In evidence theory, a **mass function** $m : 2^\Omega \to [0, 1]$ defined over the power set of Ω represents our uncertainty about the value of θ. The mass $m(E)$ can represent various things depending on the chosen interpretation:

- amount of evidence given to the fact that E contains the true value [2, 3],
- or the frequency of the imprecise observation E [1].

Mass functions sum to one, i.e., $\sum_{E \in 2^\Omega} m(E) = 1$. A set E receiving positive mass $m(E) > 0$ is called a **focal element**. We will denote by $|A|$ the **cardinality** of a set A. In particular, $|\Omega| = n$ and $|2^\Omega| = N = 2^n$.

Several alternative set-functions can then be defined to represent the same information as the one encoded in a mass function. The main ones are the plausibility, belief, implacability and commonality functions. The **plausibility** function $pl : 2^\Omega \to [0, 1]$ is defined as

$$(1) \qquad pl(A) = \sum_{E \cap A \neq \emptyset} m(E)$$

and evaluates how much event A (being true) is consistent with the current evidence. The **belief** $bel : 2^\Omega \to [0, 1]$ and **implacability** $b : 2^\Omega \to [0, 1]$ functions are defined as

$$(2) \qquad bel(A) = \sum_{E \subseteq A, E \neq \emptyset} m(E),$$

$$(3) \qquad b(A) = \sum_{E \subseteq A} m(E) = bel(A) + m(\emptyset).$$

Both evaluate how much event A (being true) is implied by the current evidence, with the implacability assuming that $\emptyset$ can imply anything, and the belief considering $\emptyset$ as inconsistent. We have $pl(A) = 1 - b(A^c)$, A^c being the complement of A. Also, we always have $bel(A) \leq pl(A)$.

When $m(\emptyset) = 0$, we have $bel = b$, and the couple belief/plausibility can be interpreted as bounds of an ill-known probability, in the sense that they induce a non-empty set

$$\mathcal{P}(m) = \{P | bel(A) \leq P(A) \leq pl(A), \forall A \subseteq \Omega\}$$

where P are probability measures over the probability space $(\Omega, 2^\Omega)$. Also, in this case, the value $pl(A) - bel(A)$ measures the imprecision of the information contained in m. When $pl(A) = bel(A)$ for all A, the set $\mathcal{P}(m)$ contains only one probability measure. This is a fully precise situation and $m(E) > 0$ only if $|E| = 1$. When $pl(A) - bel(A) = 1$ for all A, $\mathcal{P}(m)$ is the set of all probability measures. This is a maximally imprecise situation and $m(\Omega) = 1$.

Requiring $m(\emptyset) = 0$ can therefore be seen as a consistency constraint, while allowing for $m(\emptyset) \neq 0$ means that some inconsistency is allowed. This inconsistency can have various origins: (i) a conflict between sources, (ii) the fact that the true value is not in Ω (open world assumption). We will call **normalized** those masses such that $m(\emptyset) = 0$.

Finally, the **commonality** function $q : 2^\Omega \to [0, 1]$ is defined as

$$(4) \qquad q(A) = \sum_{E \supseteq A} m(E), \; \forall A \in 2^\Omega.$$

It evaluates how much A is common, *i.e.*, how much it implies the current evidence. While interpreting the commonality is more difficult, it plays an important role in evidence theory, and is a relevant measurement of how much information m contains. Indeed, suppose m_1 and m_2 are two mass functions such that each focal element E_2 of m_2 is a superset of a focal element E_1 of m_1: $E_1 \subseteq E_2$. We have $q_1(A) \leq q_2(A)$ for any $A \in 2^\Omega$ and one can fairly consider m_2 as less informative than m_1.

Each of the functions m, pl, bel, b, q contains the same information, in the sense that there exists bijective transformations to go from one to the other [2, 20].

Two other functions computed from a mass m will be used in the sequel. The contour function [2] $\pi : \Omega \to [0, 1]$ of m is defined as

$$\pi(\omega_i) = pl(\{\omega_i\})$$

and is the trace of pl on singletons. The pignistic probability distribution[3, 21] $BetP : \Omega \to [0, 1]$ of m is defined as

$$BetP(\omega_i) = \sum_{\omega_i \in E} m(E)/|E|.$$

In contrast with the other representations, they only represent a part of the information of m, and are not in bijection with it (two distinct masses $m_1 \neq m_2$ may have the same contour functions $\pi_1 = \pi_2$ or the same pignistic probability $BetP_1 = BetP_2$).

Some particular mass functions that will be instrumental in what follows are **categorical** mass functions. A categorical mass function m_E is such that $m_E(E) = 1$, and it represents exactly the set E. A specific case of categorical mass function is the **vacuous** one m_Ω, that represents complete ignorance about the true value of θ.

Besides, mass functions can be viewed as vectors belonging to the vector space $\mathbb{R}^N$ with categorical mass functions as base vectors. Since mass functions sum to one, the set of mass functions is the simplex $\mathcal{M}$ in that vector space whose vertices are the base vectors $\{m_A\}_{A \subseteq \Omega}$. This simplex is also called **mass space** [22].

3. Interpreting distances through partial orders

This section introduces the basic idea used in the rest of the paper: interpreting distances through their compatibility with partial orders. In the sequel, we apply this idea to the most well-known and characterized partial orders, that is the one comparing informative contents of belief functions.
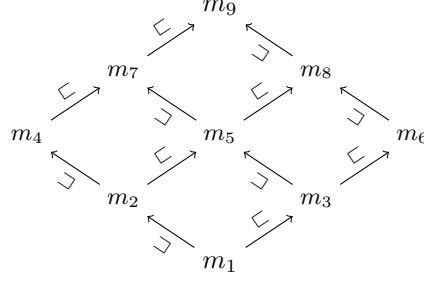
FIGURE 1. Illustration of partially ordered mass functions

3.1. **orders.** A first way to characterize the underlying structure of the mass space $\mathcal{M}$ is to look for binary relations allowing pairwise comparisons of mass functions, and in particular partial orders. A **binary relation** is a subset R of $\mathcal{M} \times \mathcal{M}$ and if a pair $(m_1, m_2) \in R$, we understand that m_1 is connected to m_2. A binary relation can thus also be seen as an oriented graph. One usually denotes alternatively $(m_1, m_2) \in R$ or $m_1 R m_2$.

A **pre-order** $\preceq$ is a binary relation over $\mathcal{M}$ such that it has the following properties:

- reflexivity: $m \preceq m$ for all $m \in \mathcal{M}$,
- transitivity: for any triplet $(m_1, m_2, m_3) \in \mathcal{M}^3$ such that $m_1 \preceq m_2$ and $m_2 \preceq m_3$, we have $m_1 \preceq m_3$.

A **partial order** $\sqsubseteq$ is a pre-order with the antisymmetry property:

- for any pair (m_1, m_2) such that $m_1 \sqsubseteq m_2$ and $m_2 \sqsubseteq m_1$, we have $m_1 = m_2$.

The pair $(\mathcal{M}, \sqsubseteq)$ is called partially ordered set or **poset** for short. If in addition, each pair (m_1, m_2) in $\mathcal{M}$ is comparable, *i.e.* we have either $m_1 \sqsubseteq m_2$ or $m_2 \sqsubseteq m_1$, then $\sqsubseteq$ is a **total order**.

Finally, an element m such that $m \sqsubseteq m_1$ and $m \sqsubseteq m_2$ is called a lower bound of $\{m_1; m_2\}$. Conversely, an element m such that $m_1 \sqsubseteq m$ and $m_2 \sqsubseteq m$ is called an upper bound of $\{m_1; m_2\}$. The **bottom** $\perp$ of a poset (if it exists) is the only element that is a lower bound of any subset of $\mathcal{M}$. The **top** $\top$ of a poset (if it exists) is the only element that is an upper bound of any subset of $\mathcal{M}$. A poset such that it is possible to find a least upper bound and a greatest lower bound for any pair (m_1, m_2) is called a **lattice**. Figure 1 illustrates a partial order between mass functions that is also a lattice.

The interest of using such qualitative relations between belief functions is that it is easier to associate them with a clear semantic. Indeed, $\sqsubseteq$ or R can be weak structures, hence can be based on simple and readable assumptions. The counterpart is that the chosen partial order $\sqsubseteq$ may contain many incomparabilities, therefore many problems relying on it will have many, possibly hard to find, solutions.

3.2. **Distances.** Another way to compare mass functions is to measure how distant m_1 and m_2 are. A **distance**, or metric, between two masses m_1, m_2 is a bounded operator $d : \mathcal{M} \times \mathcal{M} \to [0, a]$ with $a \in \mathbb{R}^+$ that satisfies the following properties:

(1) Symmetry : $d(m_1, m_2) = d(m_2, m_1)$,
(2) Definiteness : $d(m_1, m_2) = 0 \Leftrightarrow m_1 = m_2$,
(3) Triangle inequality : $d(m_1, m_2) \leq d(m_1, m_3) + d(m_3, m_2), \forall m_3 \in \mathcal{M}$.

A **pseudo distance** is an operator that satisfies symmetry and triangle inequality, but not definiteness (two distinct masses may have a zero pseudo distance value).

The pair $(\mathcal{M}, d)$ is a called a **metric space**. When the mass space $\mathcal{M}$ is endowed with a distance d, then d is called an **evidential distance**.

Jousselme and Maupin [16] have reviewed a large spectrum of operators within evidence theory associated to distances, checking among other things which properties (among Symmetry, Definiteness and Triangle inequality) were satisfied by each. In this paper, we will only study proper distances and pseudo-distances, therefore retaining those operators having the nicest mathematical properties. More recently, some authors have studied the consistency or the links of such distances with other notions of evidence theory. For instance, Loudahi *et al.* [17, 18] studies the consistency of distances with information fusion operators, such as the conjunctive rule and some of its extensions. In some sense, this paper pursues this effort by building and investigating links between the mass space seen as a poset (induced by a partial order) and the mass space seen as a metric space.

Distances require to impose a much richer structure on $\mathcal{M}$ than partial orders or binary relations, as they induce a numerical (and therefore totally ordered) relation between any pair of mass functions. On one hand, this allows to perform much more sophisticated numerical operations on mass functions, but on the other hand the relation they induce on mass functions is much harder to interpret in general, especially as $\mathcal{M}$ is a large, complicated space.

3.3. **Compatibility between partial orders and distances.** In our opinion, one way to combine the advantages of partial orders or binary relations (potentially high interpretability) and of distances (rich structure on $\mathcal{M}$ and access to numerical tools) is to interpret distances through their compatibility or consistency with partial orders. Formally, this can be defined as follows:

Definition 1. Given a partial order $\sqsubseteq_y$ defined over $\mathcal{M}$, An evidential (pseudo) distance d is said to be $\sqsubseteq_y$-**compatible** if for any mass functions m_1, m_2 and m_3 such that $m_1 \sqsubseteq_y m_2 \sqsubseteq_y m_3$, we have:

$$(5) \qquad \max \{d(m_1, m_2) ; d(m_2, m_3)\} \leq d(m_1, m_3),$$

Moreover, d is said to be $\sqsubset_y$-**compatible (in the strict sense)** if $m_1 \sqsubset_y m_2 \sqsubset_y m_3$ implies a strict inequality: $\max \{d(m_1, m_2) ; d(m_2, m_3)\} < d(m_1, m_3)$.

In particular, if the partial order $\sqsubseteq_y$ has[1] a bottom $\bot \in \mathcal{M}$ and a top element $\top \in \mathcal{M}$, then satisfying strict compatibility in Definition 1 ensures that d refines the partial order $\sqsubseteq_y$ into a total pre-order $\preceq_y$ defined as $m_1 \preceq_y m_2$ if $d(\bot, m_1) \leq d(\bot, m_2)$.

Conversely, we will say that a distance is not compatible, or **incompatible**, with a partial order if Definition 1 is not satisfied for some triplet m_1, m_2, m_3, that is $m_1 \sqsubseteq_y m_2 \sqsubseteq_y m_3$ and $\max \{d(m_1, m_2) ; d(m_2, m_3)\} > d(m_1, m_3)$.

The trivial distance[2] is obviously compatible with any non-strict partial order and incompatible with any strict order. In general, this tends to show that $\sqsubset_y$-compatibility in the strict sense is a lot more valuable property than $\sqsubseteq_y$-compatibility.

When possible, we can use implications between different orders (see Eq. (6)) to avoid checking the consistency of a distance with respect to all partial orders, as shown by the next proposition.

Proposition 1. *Consider two partial orders $\sqsubseteq_x$, $\sqsubseteq_y$ such that $\sqsubseteq_x \Rightarrow \sqsubseteq_y$ and a distance d, then*

---

[1] Recall that bottom and top elements $\bot$, $\top$ of $\sqsubseteq$ are such that for any other element x, $\bot \sqsubset x \sqsubset \top$.

[2] For any m and m', the trivial metric equals 1 whenever $m \neq m'$.

- *if d is $\sqsubseteq_y$-compatible, then it is $\sqsubseteq_x$-compatible;*
- *if d is $\sqsubseteq_x$-incompatible, then it is $\sqsubseteq_y$-incompatible.*

Proof. The second implications follows from the first by contraposition, hence we will only show the first. For this, take any triplet m_1, m_2, m_3 such that $m_1 \sqsubseteq_x m_2 \sqsubseteq_x m_3$. We then have

$$m_1 \sqsubseteq_x m_2 \sqsubseteq_x m_3 \Rightarrow m_1 \sqsubseteq_y m_2 \sqsubseteq_y m_3$$
$$\Rightarrow \max\{d_{12}; d_{23}\} \leq d_{13}$$

where $d_{ij} = d(m_i, m_j)$. The first implication following from $\sqsubseteq_x \Rightarrow \sqsubseteq_y$, and the second from the $\sqsubseteq_y$-compatibility of d.   □

An immediate corollary follows concerning the strict part:

Corollary 1. *Consider two partial strict orders $\sqsubset_x$, $\sqsubset_y$ such that $\sqsubset_x \Rightarrow \sqsubset_y$ and a distance d, then*

- *if d is strictly $\sqsubset_y$-compatible, then it is strictly $\sqsubset_x$-compatible;*
- *if d is strictly $\sqsubset_x$-incompatible, then it is strictly $\sqsubset_y$-incompatible.*

We can now apply this idea to the most studied partial orders in belief functions, that is the one comparing informative contents, and study from it those distances that are the most adapted to solve the important problem of approximating complex belief functions by simpler ones.

4. Approximating belief functions through distances

We first recall the partial orders based on informative content, before proceeding to the study of distances that can be used to approximate belief functions.

4.1. **Comparing Informative content.** Partial orders comparing informative contents formalise the notion of inclusion between belief functions, and play an essential role in approximations problem, as it is usual to require the approximation to be either an inner approximation included in the initial belief functions, or a conservative outer approximation including the initial belief function. To define such inclusions, we then have to rely on those orders. Several definitions are found in the literature:

i) m_1 is **pl-included** in m_2, denoted $m_1 \sqsubseteq_{pl} m_2$, if $pl_1(A) \leq pl_2(A)$ for all $A \in 2^\Omega$, where pl_i is the plausibility induced by m_i.

ii) m_1 is **b-included** in m_2, denoted $m_1 \sqsubseteq_b m_2$, if $b_1(A) \geq b_2(A)$ for all $A \in 2^\Omega$, where b_i is the implicability induced by m_i.

iii) m_1 is **bel-included** in m_2, denoted $m_1 \sqsubseteq_{bel} m_2$, if $bel_1(A) \geq bel_2(A)$ for all $A \in 2^\Omega$, where bel_i is the belief induced by m_i.

iv) m_1 is **q-included** in m_2, denoted $m_1 \sqsubseteq_q m_2$, if $q_1(A) \leq q_2(A)$ for all $A \in 2^\Omega$, where q_i is the commonality induced by m_i.

v) m_1 is **π-included** in m_2, denoted $m_1 \preceq_\pi m_2$, if $\pi_1(\omega) \leq \pi_2(\omega)$ for all $\omega \in \Omega$, where π_i is the contour function induced by m_i.

vi) A function m_1 is a **specialization** of m_2, denoted $m_1 \sqsubseteq_s m_2$, if there exists a non-negative $N \times N$ matrix $\mathbf{S} = [S(i,j)]$ such that

$$\text{for } j = 1, \ldots, N, \quad \sum_{i=1}^{N} S(i,j) = 1,$$

$$S(i,j) > 0 \Rightarrow E_i \subseteq E'_j,$$

$$\text{for } i = 1, \ldots, N, \quad \sum_{j=1}^{N} m_2(E'_j) S(i,j) = m_1(E_i).$$

The term $S\,(i,j) > 0$ is the proportion of the focal set E'_j that "flows down" to focal set E_i. The order in which subsets are indexed is arbitrary.

The strict version $\sqsubset$ of these inclusions is simply obtained when the inequalities are strict for at least one element, and in the case of s-inclusion when $S_{ij} > 0$ for at least one pair $E_i \subseteq E'_j$.

All these concepts extend classical set inclusion, in the sense that if $A \subseteq B$, then $m_A \sqsubseteq_x m_B$ for any $x \in \{pl, b, bel, q, s, \pi\}$. It is well known that set-inclusion is a partial order over 2^Ω. Likewise, these binary relations are partial orders over $\mathcal{M}$ except $\preceq_\pi$ which is just a pre-order. These partial orders are not total orders in the sense that we may have $m_1 \not\sqsubseteq_x m_2$ and $m_2 \not\sqsubseteq_x m_1$. They induce a lattice structure over the mass space (whose top is m_Ω and bottom is $m_\emptyset$). Again, there exists other orders extending set inclusion [23], but these are the most canonical ones.

Due to the duality between pl and b, pl- and b-inclusions are equivalent notions. When $m(\emptyset) = 0$, then pl-, b- and bel-inclusion coincide, and $m_1 \sqsubseteq_{pl} m_2$ is then equivalent to $\mathcal{P}(m_1) \subseteq \mathcal{P}(m_2)$.

The following implications hold between these notions of inclusion [24]:

$$(6) \qquad m_1 \sqsubseteq_s m_2 \Rightarrow \left\{ \begin{array}{c} m_1 \sqsubseteq_{pl} m_2 \\ m_1 \sqsubseteq_q m_2 \end{array} \right\} \Rightarrow m_1 \preceq_\pi m_2.$$

Since notions of pl, q and s-inclusion are antisymmetric, that is $m_1 \sqsubseteq_x m_2$ and $m_1 \sqsupseteq_x m_2$ implies $m_1 = m_2$ for $x \in \{pl, q, s\}$, we also have

$$(7) \qquad m_1 \sqsubset_s m_2 \Rightarrow \left\{ \begin{array}{l} m_1 \sqsubset_{pl} m_2 \\ m_1 \sqsubset_q m_2 \end{array} \right.$$

Due to Proposition 1, it is then sufficient to show that a distance is $\sqsubset_{pl} -$ or $\sqsubset_q -$compatible to show that it is $\sqsubset_s -$compatible, or that it is $\sqsubset_s -$incompatible to show incompatibility with the other partial orders. The same holds for any orders that implies $\sqsubset_s -$inclusion (e.g., Dempsterian specialization [25], orders induced by conjunctive or disjunctive weights [23]).

4.2. **Distances compatible with informative content orders.** To introduce the distances studied in this paper, we must first recall that functions m, pl, b, bel, q, π can be encoded as vectors. In general, it is sufficient to choose an arbitrary way to index subsets of Ω: $2^\Omega = (E_i)_{1 \le i \le N}$. However, in practice, a convenient way to index subsets is given by the binary order [26]. The details of this representation are recalled in appendix A.

For a given set-function $x \in \{m, pl, b, bel, q\}$, we will use a bold letter $\mathbf{x}$ to denote the corresponding vector of N elements such that the ith element of $\mathbf{x}$ is $x(E_i)$. Similarly, we will denote by $\boldsymbol{\pi}$ and $\mathbf{BetP}$ the vector of n elements whose i^{th} element is respectively $\pi(\omega_i)$ and $BetP(\omega_i)$.

In evidence theory, the most widely used metric is Jousselme distance [27]. It is based on an inner product relying on a similarity matrix. This distance is given by:

$$(8) \qquad d_J(m_1, m_2) = \sqrt{\frac{1}{2}(\mathbf{m_1} - \mathbf{m_2})' \, \mathbf{Jac}\, (\mathbf{m_1} - \mathbf{m_2})},$$

with $\mathbf{Jac}$ the Jaccard similarity matrix between sets. Its components are:

$$(9) \qquad Jac(i,j) = \begin{cases} 1 & \text{if } E_i = E_j = \emptyset \\ \frac{|E_i \cap E_j|}{|E_i \cup E_j|} & \text{otherwise} \end{cases}.$$

The success of Jousselme distance is explained by the fact that, thanks to the matrix $\mathbf{Jac}$, the relations between focal elements are taken in account. Consequently, the poset structure of 2^Ω has an impact on distance values, allowing a better match with the user's expectations.

Many other evidential distances are defined similarly by substituting matrix **Jac** with another matrix evaluating the similarity between base vectors in different ways [28, 22]. Some experimental results [27] show that these distances are highly correlated to d_J.

Another family of distances is covered by considering norms over vectors, as any norm induces a distance defined as the norm of the difference between a pair of elements. In this paper, we focus on L^k norms whose definition is:

$$
(10) \qquad \|\mathbf{x}\|_k = \left(\sum_{1 \leq i \leq p} |\mathbf{x}_i|^k \right)^{\frac{1}{k}},
$$

with $\mathbf{x}$ a given vector of size $p = N$ or n. For each category $x \in \{m; bel; pl; q; b; \pi; BetP\}$, several normalized evidential distances can be defined as follows:

$$
\begin{aligned}
d_{x,k} : \mathcal{M} \times \mathcal{M} &\rightarrow [0,1], \\
(11) \qquad m_1 \times m_2 &\rightarrow \frac{1}{a} \|\mathbf{x}_1 - \mathbf{x}_2\|_k,
\end{aligned}
$$

with $\|.\|_k$ the L^k norm, $k \in \mathbb{N}^*$, the set of positive natural numbers, and $\mathbf{x}_i$ the vector containing the values of the function x_i induced by m_i, $i \in \{1; 2\}$. If $x \in \{\pi; BetP\}$, $d_{x,k}$ is a pseudo-metric whereas if $x \in \{bel; pl; q; b\}$, $d_{x,k}$ is a metric. From [18], the normalizing constant a is given by:

$$
(12) \qquad a = \max_{A,B \in 2^\Omega} d_{x,k} (m_A, m_B).
$$

Distances induced by L^k norms have already been used in evidence theory. In their survey (section 3.2), Jousselme and Maupin[16] recall the history of such distances, also known as the Minkowski family of evidential distances. As already remarked (but not formalised) in their paper, one can see that some of the distances (11) are equivalent. In particular:

Lemma 1. *for any $k \in \mathbb{N}^*$, we have:*

$$
(13) \qquad d_{b,k} \quad = \quad d_{pl,k}.
$$

Proof. Let m_1 and m_2 be two mass functions on 2^Ω. By definition of the L^k norm and the distance $d_{b,k}$, for any positive finite integer k, we have :

$$
\begin{aligned}
\|\mathbf{b}_1 - \mathbf{b}_2\|_k &= \left(\sum_{A \subseteq \Omega} |b_1(A) - b_2(A)|^k \right)^{\frac{1}{k}}, \\
&= \left(\sum_{A \subseteq \Omega} |1 - pl_1(A^c) - 1 + pl_2(A^c)|^k \right)^{\frac{1}{k}}, \\
&= \left(\sum_{A \subseteq \Omega} |pl_1(A) - pl_2(A)|^k \right)^{\frac{1}{k}}, \\
&= \|\mathbf{pl}_1 - \mathbf{pl}_2\|_k,
\end{aligned}
$$

The last statement is equivalent to $d_{b,k}(m_1, m_2) = d_{pl,k}(m_1, m_2)$. If $k = \infty$, the same reasoning applies:

$$
\begin{aligned}
\|\mathbf{b}_1 - \mathbf{b}_2\|_\infty &= \max_{A \subseteq \Omega} |b_1(A) - b_2(A)|, \\
&= \max_{A \subseteq \Omega} |1 - pl_1(A^c) - 1 + pl_2(A^c)|, \\
&= \max_{A \subseteq \Omega} |pl_1(A) - pl_2(A)|, \\
&= \|\mathbf{pl}_1 - \mathbf{pl}_2\|_\infty
\end{aligned}
$$

and thus $d_{b,\infty} = d_{pl,\infty}$. $\qquad\square$

An immediate corollary is the following

Corollary 2. *for any $k \in \mathbb{N}^*$ and any normalized mass functions m_1 and $m_2 \in \mathcal{M}$ we have:*

$$
d_{b,k}(m_1, m_2) = d_{bel,k}(m_1, m_2) = d_{pl,k}(m_1, m_2),
$$

(14)

These preliminary results allow to simplify some upcoming proofs. We are now ready to study which of the previously mentioned distances are compatible with partial order introduced in Section 4.1.

First note that Loudahi *et al.* [18] already introduced a concept similar to ours focused on sets (or categorical mass functions), even if they did discuss it with a different purpose in mind (*i.e.*, formalizing Jousselme and Maupin [16] idea that a distance should reflect interactions between focal sets). This property is called ⊆**-compatibility**. In this article, we give a slightly different definition of this property which reads as follows:

Definition 2. An evidential distance d is said to be ⊆**-compatible** if its restriction to categorical mass functions is not the trivial set distance and if $\forall A, B, C \subseteq \Omega$ such that $A \subseteq B \subseteq C$, one has:

(15)
$$
\max\{d(m_A, m_B); d(m_B, m_C)\} \le d(m_A, m_C).
$$

This property ensures that evidential distances between categorical mass functions have dynamics similar to relevant set distances such as the Jaccard distance or the Hamming set distance. In [18], the term $d(m_B, m_C)$ is not taken into account in the inequality, which boils down to :

(16)
$$
d(m_A, m_B) \le d(m_A, m_C).
$$

Definition 1 can obviously be seen as a straightforward generalization of Definition 2 provided that the order $\sqsubseteq_y$ is such that $A \subseteq B \Rightarrow m_A \sqsubseteq_y m_B$. This assertion holds for $y \in \{pl, q, bel, s, \pi\}$, as already mentioned in [29, Sec. 2.3.]. The following Lemma follows:

Lemma 2. *Any $\sqsubseteq_y$-compatible evidential distance with $y \in \{pl, q, bel, s, \pi\}$ is also ⊆-compatible.*

Proof. From Destercke and Dubois [29, Sec. 2.3.], we have that $A \subseteq B$ implies $m_A \sqsubseteq_y m_B$ for $y \in \{pl, q, bel, s, \pi\}$, that is

$$
A \subset B \Rightarrow m_A \sqsubseteq_s m_B \Rightarrow \left\{ \begin{array}{c} m_A \sqsubseteq_{pl} m_B \\ m_A \sqsubseteq_q m_B \end{array} \right\} \Rightarrow m_A \preceq_\pi m_B.
$$

Now, if a distance is $\sqsubseteq_y$-compatible, it is sufficient to apply the first implication of Proposition 1 to see that it is also ⊆-compatible. $\qquad\square$

Proposition 2. *For any $k \in \mathbb{N}^* \setminus \{\infty\}$, the following assertions hold:*

- *the distances $d_{b,k}$ and $d_{pl,k}$ are $\sqsubset_{pl}$ and $\prec_\pi$-compatible in the strict sense,*
- *the distance $d_{bel,k}$ is $\sqsubset_{bel}$-compatible in the strict sense,*
- *the distance $d_{q,k}$ is $\sqsubset_q$ and $\prec_\pi$-compatible in the strict sense.*
- *the pseudo-distance $d_{\pi,k}$ is $\prec_\pi$-compatible in the strict sense.*

The same results also hold for $k = \infty$ with non-strict orders.

Proof. We will start with distance $d_{pl,k}$ with $k < \infty$, and will then proceed with the others:

- for distances $d_{pl,k}$: let pl_1, pl_2 and pl_3 denote three plausibility functions induced by m_1, m_2, m_3. Let us suppose that $m_1 \sqsubset_{pl} m_2 \sqsubset_{pl} m_3$. We can write $d_{pl,k}(m_1, m_3)$

$$
\begin{aligned}
&= \left(\|\mathbf{pl}_1 - \mathbf{pl}_3\|_k \right)^k, \\
&= \sum_{A \subseteq \Omega} |pl_1(A) - pl_3(A)|^k.
\end{aligned}
$$

Since $m_1 \sqsubset_{pl} m_2 \sqsubset_{pl} m_3$, we know that for any $A \subseteq \Omega$:

$$|pl_1(A) - pl_3(A)| \geq \max\{pl_3(A) - pl_2(A); pl_2(A) - pl_1(A)\} \geq 0$$

and that the inequality is strict for at least one subset. This gives :

$$
d_{pl,k}(m_1, m_3) > \sum_{A \subseteq \Omega} \max\left\{ (pl_3(A) - pl_2(A))^k, \right.
$$
$$
\left. (pl_2(A) - pl_1(A))^k \right\}
$$

As the sum of maxima is always higher than the maximum of the sums, this gives

$$
d_{pl,k}(m_1, m_3) > \max\left\{ \sum_{A \subseteq \Omega} (pl_3(A) - pl_2(A))^k, \right.
$$
$$
\left. \sum_{A \subseteq \Omega} (pl_2(A) - pl_1(A))^k \right\}
$$
$$
> \max\left\{ \left(\|\mathbf{pl}_1 - \mathbf{pl}_2\|_k \right)^k, \right.
$$
$$
\left. \left(\|\mathbf{pl}_2 - \mathbf{pl}_3\|_k \right)^k \right\}.
$$

The last inequality is equivalent to :

$$
(17) \qquad d_{pl,k}(m_1, m_3) > \max\left\{ d_{pl,k}(m_1, m_2); d_{pl,k}(m_2, m_3) \right\}.
$$

- for distances $d_{b,k}$: given equation (17) and Lemma 1, the proof is immediate.
- for distances $d_{bel,k}$: the proof is actually similar to the one for $d_{pl,k}$. Let bel_1, bel_2 and bel_3 denote three belief functions induced by m_1, m_2, m_3 and suppose that $m_1 \sqsubset_{bel} m_2 \sqsubset_{bel} m_3$. We then have:

$$
\begin{aligned}
d_{bel,k}(m_1, m_3)^k &= \left(\|\mathbf{bel}_1 - \mathbf{bel}_3\|_k \right)^k, \\
&= \sum_{A \subseteq \Omega} |bel_1(A) - bel_3(A)|^k.
\end{aligned}
$$

Since $m_1 \sqsubset_{bel} m_2 \sqsubset_{bel} m_3$, we know that for any $A \subseteq \Omega$:

$$|bel_1(A) - bel_3(A)| \geq \max\{bel_2(A) - bel_3(A); bel_1(A) - bel_2(A)\} \geq 0$$

and that the inequality is strict for at least one subset. The proof then follows by the same reasoning as for plausibilities.

- for distances $d_{q,k}$, the proof follows the same pattern. Simply consider q_1, q_2, q_3 induced by m_1, m_2, m_3 such that $m_1 \sqsubseteq_q m_2 \sqsubseteq_q m_3$, then

$$
\begin{aligned}
d_{q,k}(m_1, m_3)^k &= \left(\| \mathbf{q}_1 - \mathbf{q}_3 \|_k \right)^k, \\
&= \sum_{A \subseteq \Omega} | q_1(A) - q_3(A) |^k.
\end{aligned}
$$

   and the proof follows similarly to the to previous cases.
- for pseudo-distances $d_{\pi,k}$, the proof follows again the same pattern (with a sum over $\omega \in \Omega$).

A complement of this proof when $k = \infty$ is given in appendix B. $\qquad\square$

We can easily show that the distances based on the pignistic probability are not compatible with the partial order $\sqsubseteq_s$, and therefore are also not compatible for any other partial order comparing informative content. Indeed, consider the following example.

Example 1. Let $\Omega = \{a, b, c\}$ and consider the mass functions

$$
m_1(\{a\}) = 1/3, \quad m_1(\{b; c\}) = 2/3
$$
$$
m_2(\{b, c\}) = 0.1, \quad m_2(\Omega) = 0.9
$$
$$
m_3(\Omega) = 1.
$$

We have $m_1 \sqsubseteq_s m_2 \sqsubseteq_s m_3$, but $BetP_1 = BetP_3$ and $BetP_2 \neq BetP_1$, hence for any L_k distance using pignistic probability, we have $d_{BetP,k}(m_1, m_3) = 0$, but $d_{BetP,k}(m_1, m_2) \neq 0$.

This shows that using $d_{BetP,k}$ is ill-advised in problems involving informative content, such as the approximation of belief functions.

A more surprising fact is that Jousselme distance d_J is unfortunately incompatible with $\sqsubseteq_y$ for $y \in \{s, q, pl, bel\}$, as show the next two counter-examples.

Example 2. Let $\Omega = \{a, b\}$ and consider the mass function vectors:

$$
\begin{aligned}
\mathbf{m}_1 &= {}^t \begin{pmatrix} 0.61 & 0.19 & 0.15 & 0.05 \end{pmatrix}, \\
\mathbf{m}_2 &= {}^t \begin{pmatrix} 0 & 0 & 0.76 & 0.24 \end{pmatrix}, \\
\mathbf{m}_3 &= {}^t \begin{pmatrix} 0 & 0 & 0.4 & 0.6 \end{pmatrix}.
\end{aligned}
$$

Let $\mathbf{S}_1$ and $\mathbf{S}_2$ denote the following specialization matrices :

$$
\mathbf{S}_1 = \begin{pmatrix} 1 & 0.6 & 0 & 0 \\ 0 & 0.4 & 0 & 0 \\ 0 & 0 & 1 & 0.6 \\ 0 & 0 & 0 & 0.4 \end{pmatrix}, \mathbf{S}_2 = \begin{pmatrix} 1 & 0 & 0.8 & 0 \\ 0 & 1 & 0 & 0.8 \\ 0 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0.2 \end{pmatrix}.
$$

We have $\mathbf{m}_1 = \mathbf{S}_1 \mathbf{m}_2 = \mathbf{S}_1 \mathbf{S}_2 \mathbf{m}_3$ which implies $m_1 \sqsubseteq_s m_2 \sqsubseteq_s m_3$, but $d_J(m_1, m_3) \approx 0.63 < d_J(m_1, m_2) \approx 0.68$, hence d_J is $\sqsubseteq_s$-incompatible. Proposition 1 gives the incompatibility with $\sqsubseteq_{pl}$ and $\sqsubseteq_q$.

Concerning $\sqsubseteq_{bel}$, another counter-example is required:

Example 3.

subset	$\emptyset$	$\{a\}$	$\{b\}$	Ω
m_1	0	0.3	0.4	0.3
m_2	0	0	0.2	0.8
m_3	0.4	0	0.1	0.5
bel_1	0	0.3	0.4	1
bel_2	0	0	0.2	1
bel_3	0	0	0.1	0.6

We have $m_1 \sqsubseteq_{bel} m_2 \sqsubseteq_{bel} m_3$, but $d_J(m_1, m_3) \approx 0.36 < d_J(m_1, m_2) \approx 0.38$, hence d_J is $\sqsubseteq_{bel}$-incompatible.

Note that substituting **Jac** in Equation (8) by other matrices may result in compatible distances, yet we leave such a systematic and tedious study to the reader interested in such more exotic distances.

We now know which distances among the one we have considered can be used to deal with the approximation problem. Proposition 2 tells us that L_k norms are good candidates (as they are compatible with the partial order derived from the measure used in the norm), while we provided counter-examples showing that other well-known distances (Jousselme distance, pignistic-based distances) ar not compatible with informative content comparisons.

4.3. **Belief function approximation.** Approximating a complex mass function by a simpler mass function "similar to it" is a typical task where distances are instrumental (*i.e.*, to define what is mathematically meant by similar). Typically, one also wants to ensure that this approximation is either an inner or an outer approximation (*e.g.*, to guarantee the cautiousness of the approximation, or to bracket the initial mass between two), according to some partial order comparing informative contents.

The following notions have been formalized by Dubois and Prade [30]. Consider a subset $\mathcal{S} \subseteq \mathcal{M}$ of target mass functions, as well as a partial order $\sqsubseteq_y$, then we have the following definition of optimal approximations:

Definition 3. A **minimal outer approximation** m_*^+ of a mass function m with respect to a partial order $\sqsubseteq_y$ within a part $\mathcal{S} \subseteq \mathcal{M}$ is a mass function such that:

(i) $m \sqsubseteq_y m_*^+$ and for any $m' \in \mathcal{S}$ s.t. $m \sqsubseteq_y m'$, $m' \not\sqsubseteq_y m_*^+$
(ii) $m_*^+ \in \mathcal{S}$.

Definition 4. A **maximal inner approximation** m_*^- of a mass function m with respect to a partial order $\sqsubseteq_y$ within a part $\mathcal{S} \subseteq \mathcal{M}$ is a mass function such that:

(i) $m_*^- \sqsubseteq_y m$ and for any $m' \in \mathcal{S}$ s.t. $m' \sqsubseteq_y m$, $m_*^- \not\sqsubseteq_y m'$,
(ii) $m_*^- \in \mathcal{S}$.

The conditions ensure that the approximation is optimal, in the sense that it is a minimal (resp. maximal) element of the outer (resp. inner) approximations of m within $\mathcal{S}$ and w.r.t. $\sqsubseteq_y$. While these conditions are quite natural, there are in general many approximations satisfying them, thus not allowing one to retrieve a unique solution. Furthermore, Definitions 3 and 4 do not provide any practical clue as to how some solutions fitting their conditions can be found. As we shall see, this can be done by using appropriate distances.

Outer and inner approximations of belief function has already been studied as a distance minimization problem by Cuzzolin [31, 15]. In these articles, the author presents some justified methodologies and geometric interpretations for particular classes of problems. We investigate these problems with a more general perspective, using distance minimization as a generic tool to find approximations. The next Lemmas show that using distances compatible with partial orders provides practical ways to get optimal outer/inner approximation.

Lemma 3. *Let m be a mass function to approximate, $\mathcal{S}$ a set of target mass functions, $\mathcal{S}_m^+$ a non-empty subset of $\mathcal{S}$ containing mass functions outer-approximating m w.r.t. a partial order $\sqsubseteq_y$, and d a $\sqsubseteq_y$-compatible distance. Then, any solution*

$$(18) \qquad m_*^+ \in \underset{m' \in \mathcal{S}_m^+}{\arg\min}\, d\,(m, m')\,.$$

is a minimal outer approximation.

Proof. By definition, m_*^+ is an outer approximation of m. Let us assume ex-absurdo that it is not minimal, meaning that there exists $m' \in \mathcal{S}_m^+$ such that

$$(19) \qquad m \sqsubseteq_y m' \sqsubset_y m_*^+.$$

However, this would mean that $d(m, m') < d(m, m_*^+)$, a contradiction since this would mean that m_*^+ is not a solution of Equation (18). $\qquad \square$

Lemma 4. *Let m be a mass function to approximate, $\mathcal{S}$ a set of target mass functions, $\mathcal{S}_m^-$ a non-empty subset of $\mathcal{S}$ containing mass functions inner-approximating m w.r.t. a partial order $\sqsubseteq_y$, and d a $\sqsubseteq_y$-compatible distance. Then, any solution*

$$(20) \qquad m_*^- \in \underset{m' \in \mathcal{S}_m^-}{\arg\min} \, d(m, m').$$

is a maximal inner approximation.

Proof. The proof is similar to the one of Lemma 3. $\qquad \square$

Note that a simple way to ensure that $\mathcal{S}_m^+$ or $\mathcal{S}_m^-$ are non-empty is to include the top and bottom (usually, the vacuous and the empty mass functions) of $\sqsubseteq_y$ in $\mathcal{S}$ when they exist. In practice, however, there may be multiple solutions to Equations (18) and (20). Since minimizing a strictly convex function over a convex set has a unique solution, the following claims hold:

- If $\mathcal{S}_m^+$ is convex and if d is induced by a strictly convex norm, then there exists only one outer approximation of m which minimizes the distance from m to the set $\mathcal{S}_m^+$.
- If $\mathcal{S}_m^-$ is convex and if d is induced by a strictly convex norm, then there exists only one inner approximation of m which minimizes the distance from m to the set $\mathcal{S}_m^-$.

Following these remarks, two corollaries are obtained.

Corollary 3. *Let m denote a mass function and $\sqsubseteq_y$ a partial order with $y \in \{pl, q, bel, s\}$. Let $\mathcal{Y}_m^+$ denote the set of all outer approximations of m with respect to $\sqsubseteq_y$. For any closed convex set $\mathcal{S} \subseteq \mathcal{M}$ such that $\mathcal{S} \cap \mathcal{Y}_m^+ \neq \emptyset$ and for any integer $1 < k < \infty$, the solution*

$$(21) \qquad m_*^+ = \underset{m' \in \mathcal{S} \cap \mathcal{Y}_m^+}{\arg\min} \, d_{y,k}(m, m').$$

is a unique minimal outer-approximation.

Proof. For any $y \in \{pl, q, bel, s\}$ and any $1 < k < \infty$, each distance $d_{y,k}$ is induced by the L^k norm which is a strictly convex function. Furthermore, since set-intersection preserves closure and convexity, one only needs to prove the closure and the convexity of $\mathcal{Y}_m^+$ for each $y \in \{pl, q, bel, s\}$. All such sets are obviously closed parts of $\mathbb{R}^N$. A proof of their convexity is given in appendix C. Now, according to convex analysis, the corollary holds. $\qquad \square$

Corollary 4. *Let m denote a mass function and $\sqsubseteq_y$ a partial order with $y \in \{pl, q, bel, s\}$. Let $\mathcal{Y}_m^-$ denote the set of all inner approximations of m with respect to $\sqsubseteq_y$. For any closed convex set $\mathcal{S} \subseteq \mathcal{M}$ such that $\mathcal{S} \cap \mathcal{Y}_m^- \neq \emptyset$ and for any integer $1 < k < \infty$, the solution*

$$(22) \qquad m_*^- = \underset{m' \in \mathcal{S} \cap \mathcal{Y}_m^-}{\arg\min} \, d_{y,k}(m, m').$$

is a unique maximal inner-approximation.

Proof. The proof is identical to that of corollary 3 except that one needs to prove the convexity of $\mathcal{Y}_m^-$ for each $y \in \{pl, q, bel, s\}$. A proof of their convexity is given in appendix C. $\qquad \square$

Corollaries 3 and 4 tell us that inner/outer optimal approximations can easily be found by solving convex optimization problems, for which many efficient algorithms (e.g., gradient descent) exist.

The main limitation of these results is that the set of constraints must define a convex subset $\mathcal{S}$ of the mass space $\mathcal{M}$. In practice, many subsets of interest will meet this requirement, for example:

- the sets of outer/inner approximations of another mass function m' with respect to any partial order $\sqsubseteq_y$ with $y \in \{pl, q, bel, s\}$;
- the set of k-additive mass functions (m is k-additive if $m(A) > 0$ only if $|A| \leq k$), including Bayesian mass functions;
- the set of mass functions whose set of focal elements is fixed (*e.g.*, possibility distribution with identical ranking of possibility degrees, convex mixture between a Bayesian and a vacuous mass);
- the set of mass functions that are a conjunctive combination between m' and any other mass function.

However, some sets of constraints are not convex. In particular, the set of consonant outer/inner approximations is notoriously not convex [32]. Yet, the results presented in this section cover a significantly wide spectrum of applications and is thus paving the way for many future works. We give a first illustrative example for the approximation of a mass function when $\mathcal{S}$ is a face of $\mathcal{M}$.

4.4. **Approximation in simplicial faces.** Approximating belief functions in a simplicial face consists in finding the closest mass function inside a face[3] of $\mathcal{M}$ from given mass function m. This task satisfies the convexity requirements of corollaries 3 and 4 as any sub-simplex is convex. In addition, such an approximation process is easy to interpret in terms of evidence updating and it is thus appropriate to illustrate the interest of the evidential distances that are studied in this paper.

Approximating in faces has been addressed by many authors [13, 31], yet we only want to illustrate that the compatibility with an evidential partial order is a desirable property for evidential distances, not to compete with state-of-the-art approaches.

Let us look for inner approximations of the Bayesian mass function $m = \frac{1}{3}m_{\{a\}} + \frac{1}{3}m_{\{b\}} + \frac{1}{3}m_{\{c\}}$ with $\Omega = \{a, b, c\}$ in the Bayesian subspace $\mathcal{S}_{[a;b]}$. For any $m' \in \mathcal{S}_{[a;b]}$, we have $m' = \lambda m_{\{a\}} + (1-\lambda)m_{\{b\}}$ with $\lambda \in [0;1]$, therefore this subspace is a segment. We will only investigate $d_{pl,k}$ and $d_{q,k}$ as we have $d_{pl,k} = d_{b,k} = d_{bel,k}$ in this experiment. Figure 2 shows distances obtained between m and any $m' \in \mathcal{S}_{[a;b]}$ for $k \in \{1, 2, \infty\}$. This gives a clear illustration that L^1 and L^∞ norms are less adequate than L^2 to seek an inner approximation, as they provide multiple minima.

Let us now look for inner approximations of the same Bayesian mass function $m = \frac{1}{3}m_{\{a\}} + \frac{1}{3}m_{\{b\}} + \frac{1}{3}m_{\{c\}}$ in the conditional subspace $\mathcal{S}_{\{a;b\}}$. For any $m' \in \mathcal{S}_{\{a;b\}}$, we have $m' = \lambda_1 m_{\{a\}} + \lambda_2 m_{\{b\}} + (1 - \lambda_1 - \lambda_2)m_{\{a,b\}}$ with $\lambda_1 \in [0;1]$ and $\lambda_2 \in [0, 1 - \lambda_1]$, therefore this subspace is a triangular face. Figure 3 shows distances obtained between m and any $m' \in \mathcal{S}_{\{a;b\}}$ for $k = 2$ only. This figure illustrates the convex shape of the problem as well as the uniqueness of the solution. From Figure 3, the intuition is that $m_*^- = \frac{1}{2}m_{\{a\}} + \frac{1}{2}m_{\{b\}}$ for both distances. When using L^2, the convex problems can be reshaped into quadratic programming problems. Solving them with function $\mathtt{qp}$ in GNU Octave confirms this intuition. The solution is reached in 6 iterations for $d_{pl,2}$ and in 7 iterations for $d_{q,2}$. This slight difference can be explained by the fact that a steeper slope is observed for $d_{pl,2}$.

---

[3]A face $\mathcal{F}$ of a simplex $\mathcal{M}$ is the simplex obtained after deletion of some vertices. Faces are ranging from facets down to edges and to vertices.
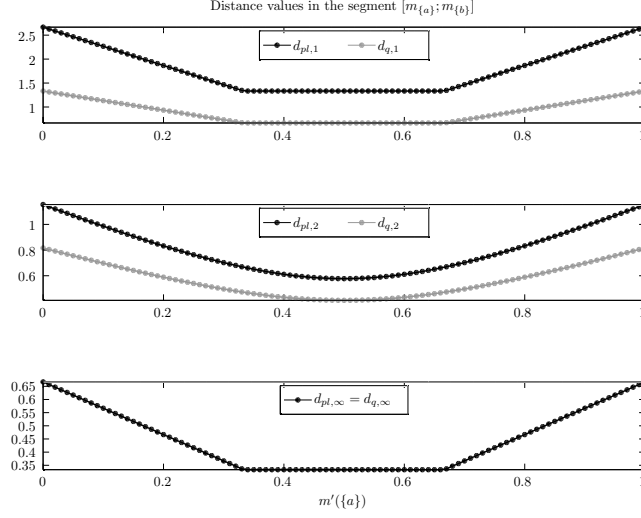
FIGURE 2. Distances $d_{pl,k}$ and $d_{q,k}$ between $m = \frac{1}{3}m_{\{a\}} + \frac{1}{3}m_{\{b\}} + \frac{1}{3}m_{\{c\}}$ and $m' = \lambda m_{\{a\}} + (1 - \lambda)m_{\{b\}}$ with $\lambda \in [0;1]$ and $k \in \{1, 2, \infty\}$. The distances are represented as functions of $m'(\{a\})$.

The term conditional subspace was coined by Cuzzolin [33]. As suggested by this terminology, the space $\mathcal{S}_{\{a;b\}}$ contains all conditional mass functions given that $\{\theta \in \{a, b\}\}$ is true. These conditional mass functions are obtained by applying Dempster's conditioning [1]. Let $m(.|A)$ denote Dempster's conditioning of m over A. We have:

$$m(X|A) = \begin{cases} \frac{1}{1-\kappa} \sum\limits_{Y = X \cap A} m(Y), & \text{if } X \in 2^\Omega \setminus \emptyset \\ 0 & \text{if } X = \emptyset \end{cases},$$

with $\kappa = \sum\limits_{Y = X \cap \emptyset} m(Y)$. From [25], we know that, in this particular situation, $m_*^- = m(.|\{a, b\})$ for $\sqsubseteq_{pl}$. Since $m(.|\{a, b\}) = \frac{1}{2}m_{\{a\}} + \frac{1}{2}m_{\{b\}}$, this result allows to validate the approximation obtained in this experiment.

5. DISCUSSION

In this section, we would like to briefly discuss a number of aspects of the current approach, as well as some possible extensions of the view presented here.

5.1. **About distance selection and interpretation.** The use of distances is now ubiquitous in works about belief functions, as recalled in the introduction. Still, the selection of an adequate distance remains a challenging task. In some cases such as classification [7, 8], some measure of performance (such as prediction accuracy) can be used to select an optimal distance, although we are not aware of any work showing the existence of a theoretical optimal distance in evidential learning methods using them.

However, in other tasks such as conflict estimation or belief function approximation, there are no such measure of performances, and in this case the choice of distance has to be based on other criteria. In addition to mathematical properties, two aspects that are important are the meaning (or semantic) of the used distance as well as its practicality.
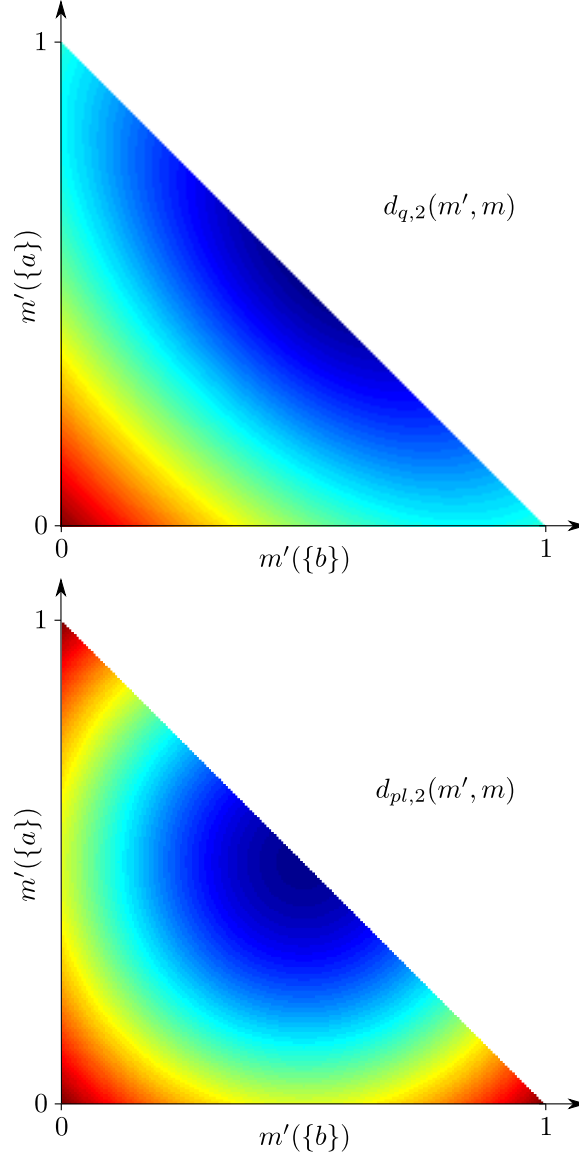
FIGURE 3. Distances $d_{pl,2}$ and $d_{q,2}$ between $m = \frac{1}{3}m_{\{a\}} + \frac{1}{3}m_{\{b\}} + \frac{1}{3}m_{\{c\}}$ and $m' = \lambda_1 m_{\{a\}} + \lambda_2 m_{\{b\}} + (1 - \lambda_1 - \lambda_2)m_{\{a,b\}}$ with $\lambda_1 \in [0;1]$ and $\lambda_2 \in [0, 1 - \lambda_1]$. The distances are represented as functions of the pair $(m'(\{a\}), m'(\{b\}))$. Colors are depicting distance values ranging from blue (lowest distance) to red (highest).

Concerning the first aspect, we have proposed a simple way to select those distances that fits a particular interpretation, by looking at their consistency with partial orders between belief functions having a well-defined semantic. Concerning the second aspect, we have shown in Section 4.3 that some distances, among the one fitted to the approximating operation, are more practical than others, in the sense that their convexity allows one to retrieve a unique, well-defined solution.

To conclude, we do not claim that consistency with partial orders should be the only way to select adequate distances, it only provides a simple way to tell whether a particular distance (or a family of distance) fits a particular semantic or not.

5.2. **About information measures.** Another common way to characterize the informative content of a belief function m is the use of information measures [20], usually extending the Shannon entropy for probabilities or the Hartley measure for sets. If we denote $I : \mathcal{M} \to \mathbb{R}$ the imprecision measure function, a common requirement is that for two masses m, m', we have

$$m \sqsupseteq_{pl} m' \Rightarrow I(m) \leq I(m).$$

See for example works of Abellan and Moral [34, 35] for extensions to more general frameworks including belief functions as special cases.

The reason why we have not considered such measures to compare informative contents is that they share the same advantages and drawbacks as (pseudo-)distances: by imposing every mass to be summarized by a value, they put a rich structure on $\mathcal{M}$ and induce a total pre-order on them. While this makes them handy numerical tools, this means that they can be hard to interpret in general situations (even if $m \not\sqsupseteq_{pl} m'$ and $m' \not\sqsupseteq_{pl} m$, $I(m)$ and $I(m')$ will still be comparable) and are likely to be consistent with only a very small set of distances (if any). They are therefore not fit to our initial purpose (providing semantics to large classes of distances).

An interesting study, nevertheless, would be to determine which distances and information measures are in accordance with each other.

5.3. **About other possible partial orders.** The orders on which we focused compare the informative contents of mass functions, yet we can think about other orders related to important notions in evidence theory. For instance, an important notion within evidence theory is the consistency of pieces of information encoded inside a mass function, from which follows the notion of conflicts between sources. There are two main ways to evaluate the consistency of a mass function, a strong and a weak one. The strong consistency measure is given by

$$(23) \qquad \Phi(m) = \max_{\omega \in \Omega} \pi(\omega)$$

and the weak consistency measure by

$$(24) \qquad \phi(m) = m(\emptyset).$$

We refer to [36] for a justification and discussion of these measures.

Each measure defines a pre-order over the mass space:

i) Mass m_1 is *strongly less consistent* than m_2, denoted $m_1 \preceq_{sc} m_2$, if $\Phi(m_1) \leq \Phi(m_2)$.

ii) Mass m_1 is *weakly less consistent* than m_2, denoted $m_1 \preceq_{wc} m_2$, if $\phi(m_1) \leq \phi(m_2)$.

Again, strict inequalities yield strict relations. $\preceq_{sc}$ and $\preceq_{wc}$ are total pre-orders, as any two elements can be compared, but distinct elements m_1, m_2 may be equally consistent. Due to this, it may be difficult to find non-trivial distances consistent with them, thereby confirming that distances may not be the most adequate tool to measure consistency [37].

This does not mean that our approach cannot be applied to issues regarding conflict and/or dependence. Indeed, one could in principle define when a pair of mass functions is "more conflicting" or "more dependent" than another pair of mass functions, for example by using interval-valued conflict measures (see [36]). A distance would then be consistent with such orders when the distance between a

pair of "more conflicting" mass functions would be lower than the distance between a pair of "less conflicting" mass functions.

6. Conclusion

We have proposed a simple way to interpret distances and to identify in which situations they can be useful, by studying their compatibility with partial orders. This compatibility is formalized into a mathematical property. We have also derived new evidential distances, justified by their compatibility with several well known partial orders in this framework. For each of the partial order considered in this paper, an infinite family of distances is proved to be compatible with respect to this given order. Each of these families rely on L^k norms, with $k \in \mathbb{N}^*$.

In addition, it is explained how such distances can be exploited as part of belief function approximation problems. In particular, minimizing such distances is equivalent to a convex optimization problem provided that the set of desired constraints defines himself a convex subset of the mass space. We show that a large number of such constraint sets are indeed convex. Consequently, there are many applicative perspectives for future work since convex optimization problem is the easiest class of optimization problem to solve.

There are also more theoretical perspectives on a more theoretical ground. For instance, one can wonder if it is possible to relate the compatibility of a distance with a partial order and its consistency with a combination rule. Another possibility would be also to define weaker forms of compatibility with pre-orders like those depicting the inner consistency of a mass function. Finally, it would be interesting to study which distances are compatible with partial orders extending stochastic dominance [38], as those are commonly used in decision problems.

Appendix A. The binary order for subset indexing

For a given finite set $\Omega = \{\omega_1, \ldots, \omega_n\}$, the representation of subsets $A \subseteq \Omega$, known as binary order, is the following: each subset A is associated to a binary number made of n bits and this number has a 1 at place i if $\omega_i \in A$, and 0 otherwise. For example, when $\Omega = \{\omega_1, \omega_2, \omega_3\}$, the binary representation of the subset $\{\omega_2, \omega_3\}$ is 110. Then, considering the integer number Int_A in base 10 obtained from the binary representation of A, each subset defines a unique vector index $n_A = Int_A + 1$ starting from 1 ($\emptyset$) to 2^n (Ω).

Appendix B. Proof of proposition 2 in the $k = \infty$ case

Proof. In the case where $k = \infty$, we will also start with $d_{pl,\infty}$ and then proceed with other distances:

- for distance $d_{pl,\infty}$: let pl_1, pl_2 and pl_3 denote three plausibility functions induced by m_1, m_2, m_3. Let us suppose that $m_1 \sqsubseteq_{pl} m_2 \sqsubseteq_{pl} m_3$. We can write:

$$
\begin{aligned}
d_{pl,\infty}(m_1, m_3) &= \|\mathbf{pl_1} - \mathbf{pl_3}\|_\infty, \\
&= \max_{A \subseteq \Omega} |pl_1(A) - pl_3(A)|.
\end{aligned}
$$

Since $m_1 \sqsubseteq_{pl} m_2 \sqsubseteq_{pl} m_3$, we know that for any $A \subseteq \Omega$, $pl_1(A) \le pl_2(A) \le pl_3(A)$ and consequently we have:

$$|pl_1(A) - pl_3(A)| \ge \max\{pl_3(A) - pl_2(A);$$

$$pl_2(A) - pl_1(A)\} \ge 0.$$

This gives :

$$d_{pl,\infty}(m_1,m_3) \geq \max_{A \subseteq \Omega} \max \left\{ pl_3(A) - pl_2(A) \,, \right.$$
$$\left. pl_2(A) - pl_1(A) \right\}$$

Note that strict relations would not imply a strict inequality in the above one. Indeed strict relations would only imply that there is at least one subset A_* with $|pl_1(A_*) - pl_3(A_*)| > \max\{pl_3(A_*) - pl_2(A_*); pl_2(A_*) - pl_1(A_*)\}$ but the maximum is not necessarily reached at that specific subset. This accounts for the fact that $d_{pl,\infty}$ is not $\sqsubseteq_{pl}$-compatible in the strict sense. The order in which the maximum is sought can be inverted, this gives

$$d_{pl,\infty}(m_1,m_3) \geq \max \left\{ \max_{X \subseteq \Omega} pl_3(X) - pl_2(X) \,, \right.$$
$$\left. \max_{X \subseteq \Omega} pl_2(X) - pl_1(X) \right\}$$
$$\geq \max \left\{ \|\mathbf{pl}_1 - \mathbf{pl}_2\|_\infty \,, \|\mathbf{pl}_2 - \mathbf{pl}_3\|_\infty \right\}.$$

The last inequality is equivalent to :

$$(25) \qquad d_{pl,\infty}(m_1,m_3) \geq \max \left\{ d_{pl,\infty}(m_1,m_2) \,; d_{pl,\infty}(m_2,m_3) \right\}$$

and $d_{pl,\infty}$ is $\sqsubseteq_{pl}$-compatible.
- for distance $d_{b,\infty}$: given equation (25) and Lemma 1, the proof is immediate.
- for distance $d_{bel,\infty}$: the proof is actually similar as for distance $d_{pl,\infty}$. Let bel_1, bel_2 and bel_3 denote three belief functions induced by m_1, m_2, m_3 and suppose that $m_1 \sqsubseteq_{bel} m_2 \sqsubseteq_{bel} m_3$. We then have:

$$d_{bel,\infty}(m_1,m_3) = \|\mathbf{bel}_1 - \mathbf{bel}_2\|_\infty \,,$$
$$= \max_{A \subseteq \Omega} |bel_1(A) - bel_3(A)|.$$

Since $m_1 \sqsubseteq_{bel} m_2 \sqsubseteq_{bel} m_3$, we know that for any $A \subseteq \Omega$:

$$|bel_1(A) - bel_3(A)| \geq \max\{bel_3(A) - bel_2(A) \,;$$
$$bel_2(A) - bel_1(A)\} \geq 0.$$

The proof then follows by the same reasoning as for plausibilities.
- for distance $d_{q,\infty}$: the proof follows the same pattern. Simply consider q_1, q_2, q_3 induced by m_1, m_2, m_3 such that $m_1 \sqsubseteq_q m_2 \sqsubseteq_q m_3$, then

$$d_{q,\infty}(m_1,m_3) = \|\mathbf{q}_1 - \mathbf{q}_3\|_\infty \,,$$
$$= \max_{X \subseteq \Omega} |q_1(X) - q_3(X)|.$$

and the proof follows similarly to the to previous cases.
- for distance $d_{\pi,\infty}$, the proof follows again the same pattern (with a sum over $\omega \in \Omega$).

$\square$

APPENDIX C. PROOF OF THE CONVEXITY OF OUTER/INNER APPROXIMATION
SETS

We begin with a proof that outer/inner approximations of a mass function m with respect to $\sqsubseteq_{pl}$ are in a convex set. We will then proceed with the other partial orders: $\sqsubseteq_{bel}$, $\sqsubseteq_q$ and $\sqsubseteq_s$. Related results are already given in [39, 40, 29, 33].

- Let $\mathcal{Pl}_m^+$ denote the set of outer approximations of m with respect to $\sqsubseteq_{pl}$. Let (m_1, m_2) be a pair of mass functions in $\mathcal{Pl}_m^+$. The mass space $\mathcal{M}$ is a simplex and is thus convex. Consequently, for any $\lambda \in [0;1]$, $m_3 = (\lambda m_1 + (1 - \lambda) m_2) \in \mathcal{M}$.
  Besides, if m_1 and m_2 are in $\mathcal{Pl}_m^+$, this means that for any $A \subseteq \Omega$, we have:

(26)
$$pl(A) \quad \leq \quad \min\{pl_1(A); pl_2(A)\}.$$

  Using equation (1), we also have:

$$\begin{aligned}
pl_3(A) &= \sum_{E \cap A \neq \emptyset} m_3(E), \\
&= \lambda \sum_{E \cap A \neq \emptyset} m_1(E) + (1 - \lambda) \sum_{E \cap A \neq \emptyset} m_2(E), \\
&= \lambda pl_1(A) + (1 - \lambda) pl_2(A), \\
&\geq \lambda \min\{pl_1(A); pl_2(A)\} + \\
&\qquad (1 - \lambda) \min\{pl_1(A); pl_2(A)\}, \\
&\geq \min\{pl_1(A); pl_2(A)\}, \\
&\geq pl(A).
\end{aligned}$$

  We thus deduce $m \sqsubseteq_{pl} m_3$ and that $\mathcal{Pl}_m^+$ is convex. The same conclusion is obtained for $\mathcal{Pl}_m^-$ by reversing the inequalities and replacing min with max.
- Let $\mathcal{Bel}_m^+$ denote the set of outer approximations of m with respect to $\sqsubseteq_{bel}$. Let (m_1, m_2) be a pair of mass functions in $\mathcal{Bel}_m^+$. For any $\lambda \in [0;1]$, $m_3 = (\lambda m_1 + (1 - \lambda) m_2) \in \mathcal{M}$.
  If m_1 and m_2 are in $\mathcal{Bel}_m^+$, this means that for any $A \subseteq \Omega$, we have:

(27)
$$bel(A) \quad \geq \quad \max\{bel_1(A); bel_2(A)\}.$$

  Similarly as in the previous case, using equation (2) and linearity, we obtain $bel_3(A) \leq \max\{bel_1(A); bel_2(A)\} \leq bel(A)$. Consequently, we deduce $m \sqsubseteq_{bel} m_3$ and that $\mathcal{Bel}_m^+$ is convex. The same conclusion is obtained for $\mathcal{Bel}_m^-$ by reversing the inequalities and replacing max with min.
- Let $\mathcal{Q}_m^+$ denote the set of outer approximations of m with respect to $\sqsubseteq_q$. Let (m_1, m_2) be a pair of mass functions in $\mathcal{Q}_m^+$. For any $\lambda \in [0;1]$, $m_3 = (\lambda m_1 + (1 - \lambda) m_2) \in \mathcal{M}$.
  If m_1 and m_2 are in $\mathcal{Q}_m^+$, this means that for any $A \subseteq \Omega$, we have:

(28)
$$q(A) \quad \leq \quad \min\{q_1(A); q_2(A)\}.$$

  Similarly as in the previous case, using equation (4) and linearity, we obtain $q_3(A) \geq \min\{q_1(A); q_2(A)\} \geq q(A)$. Consequently, we deduce $m \sqsubseteq_q m_3$ and that $\mathcal{Q}_m^+$ is convex. The same conclusion is obtained for $\mathcal{Q}_m^-$ by reversing the inequalities and replacing min with max.
- Let $\mathcal{Spe}_m^+$ denote the set of outer approximations of m with respect to $\sqsubseteq_s$. Let (m_1, m_2) be a pair of mass functions in $\mathcal{Spe}_m^+$. For any $\lambda \in [0;1]$, $m_3 = (\lambda m_1 + (1 - \lambda) m_2) \in \mathcal{M}$. The cases $\lambda = 0$ and $\lambda = 1$ are trivial, we will thus consider that $\lambda \in]0;1[$.

Let $\mathbf{S}_3$ denote a $N \times N$ matrix whose elements are defined as follows:

$$(29) \qquad S_3(i,j) = \begin{cases} \delta_{ij} \text{ if } \lambda m_1(E_j) + (1-\lambda) m_2(E_j) = 0 \\ \frac{\lambda m_1(E_j)S_1(i,j)+(1-\lambda)m_2(E_j)S_2(i,j)}{\lambda m_1(E_j)+(1-\lambda)m_2(E_j)} \text{ otherwise} \end{cases},$$

with δ_{ij} the Kronecker delta.

Let us first prove that $\mathbf{S}_3$ is a specialization matrix. Obviously, $\mathbf{S}_3$ is non-negative. Let $\rho_j = m_3(E_j) = \lambda m_1(E_j) + (1-\lambda) m_2(E_j)$. For any j from 1 to N such that $\rho_j > 0$, we also have:

$$\begin{aligned} \sum_{i=1}^N S_3(i,j) &= \sum_{i=1}^N \frac{\lambda m_1(E_j)S_1(i,j)+(1-\lambda)m_2(E_j)S_2(i,j)}{\rho_j}, \\ &= \frac{1}{\rho_j}\left(\lambda m_1(E_j) \sum_{i=1}^N S_1(i,j) \right. \\ &\qquad \left. + (1-\lambda) m_2(E_j) \sum_{i=1}^N S_2(i,j)\right), \\ &= \frac{\lambda m_1(E_j) + (1-\lambda) m_2(E_j)}{\rho_j}, \\ &= 1. \end{aligned}$$

This results also holds if $\rho_j = 0$. Moreover, if $S_3(i,j) > 0$ and $\rho_j > 0$ then either $S_1(i,j) > 0$ or $S_2(i,j) > 0$. Since $\mathbf{S}_1$ and $\mathbf{S}_2$ are specialization matrices, it implies that $E_i \subseteq E_j$. If $S_3(i,j) > 0$ and $\rho_j = 0$ then $\delta_{ij} > 0 \Rightarrow E_i = E_j$. In conclusion, $\mathbf{S}_3$ is a specialization matrix.

Let us now prove that $\mathbf{m} = \mathbf{S}_3 \mathbf{m}_3$. For any i from 1 to N, we have:

$$\sum_{j=1}^N S_3(i,j) m_3(E_j)$$

$$\begin{aligned} &= \sum_{\substack{1 \leq j \leq N \\ \text{s.t. } \rho_j > 0}} S_3(i,j) \rho_j + \sum_{\substack{1 \leq j \leq N \\ \text{s.t. } \rho_j = 0}} S_3(i,j) \rho_j, \\ &= \sum_{\substack{1 \leq j \leq N \\ \text{s.t. } \rho_j > 0}} \lambda m_1(E_j)S_1(i,j)+(1-\lambda)m_2(E_j)S_2(i,j). \end{aligned}$$

When $\rho_j = 0$, we also have $m_1(E_j) = m_2(E_j) = 0$. One can thus write:

$$\sum_{j=1}^N S_3(i,j) m_3(E_j)$$

$$\begin{aligned} &= \sum_{1 \leq j \leq N} \lambda m_1(E_j)S_1(i,j)+(1-\lambda)m_2(E_j)S_2(i,j), \\ &= \lambda \sum_{1 \leq j \leq N} m_1(E_j)S_1(i,j)+ (1-\lambda) \sum_{1 \leq j \leq N} m_2(E_j)S_2(i,j), \\ &= \lambda m(E_i) + (1-\lambda) m(E_i), \\ &= m(E_i). \end{aligned}$$

This means that $m \sqsubseteq_s m_3$ and that $\mathcal{S}pe_m^+$ is convex.

Let $\mathcal{S}pe_m^-$ denote the set of inner approximations of m with respect to $\sqsubseteq_s$. Let (m_1, m_2) be a pair of mass functions in $\mathcal{S}pe_m^-$. For any $\lambda \in [0;1]$, $m_3 = (\lambda m_1 + (1-\lambda) m_2) \in \mathcal{M}$.

If m_1 and m_2 are in $\mathcal{S}pe_m^-$, this means that there exists two matrices $\mathbf{S}_1$

and $\mathbf{S}_2$ with $\mathbf{m}_1 = \mathbf{S}_1 \mathbf{m}$ and $\mathbf{S}_2 \mathbf{m} = \mathbf{m}_2$. We have:

$$
\begin{aligned}
(\lambda \mathbf{S}_1 + (1 - \lambda) \mathbf{S}_2) \mathbf{m} &= \lambda \mathbf{S}_1 \mathbf{m} + (1 - \lambda) \mathbf{S}_2 \mathbf{m}, \\
&= \lambda \mathbf{m}_1 + (1 - \lambda) \mathbf{m}_2, \\
&= \mathbf{m}_3.
\end{aligned}
$$

We deduce $m_3 \sqsubseteq_s m$ and that $\mathcal{S}pe_m^-$ is convex.

References

[1] A. Dempster, "Upper and lower probabilities induced by a multiple valued mapping," *Annals of Mathematical Statistics*, vol. 38, pp. 325 – 339, 1967.

[2] G. Shafer, *A mathematical theory of evidence*, . p. Princeton University Press, Ed. Princeton University Press, 1976.

[3] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intelligence*, vol. 66, no. 2, pp. 191 – 234, 1994.

[4] W. Liu, "Measuring conflict between possibilistic uncertain information through belief function theory," in *Knowledge Science, Engineering and Management*, ser. Lecture Notes in Computer Science, J. Lang, F. Lin, and J. Wang, Eds. Springer Berlin Heidelberg, 2006, vol. 4092, pp. 265–277.

[5] A. Martin, A.-L. Jousselme, and C. Osswald, "Conflict measure for the discounting operation on belief functions," in *Proceedings of the 11th International Conference on Information Fusion (FUSION 2008)*, Cologne, Germany, June 2008, pp. 1–8.

[6] M. Chebbah, A. Martin, and B. B. Yaghlane, "Combining partially independent belief functions," *Decision Support Systems*, vol. 73, no. 0, pp. 37 – 46, 2015.

[7] L. M. Zouhal and T. Denœux, "An evidence-theoretic k-nn rule with parameter optimisation," *IEEE Transactions on Systems, Man and Cybernetics. Part C: Application and reviews*, vol. 28, no. 2, pp. 263 – 271, 1998.

[8] D. Fixen and R. Mahler, "The modified Dempster-Shafer approach to classification," *IEEE Transactions on Systems, Man and Cybernetics. Part A: Systems and humans*, vol. 27, no. 1, pp. 96 – 104, 1997.

[9] D. Harmanec, "Faithful approximations of belief functions," in *Uncertainty in Artificial Intelligence (UAI99) 15*, K. Laskey and H. Prade, Eds., Stockholm (Sweden), 1999.

[10] M. Bauer, "Approximation algorithms and decision making in the Dempster-Shafer theory of evidence an empirical study," *International Journal of Approximate Reasoning*, vol. 17, no. 23, pp. 217 – 237, 1997.

[11] T. Weiler and U. Bodenhofer, "Approximation of belief functions by minimizing euclidean distances," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 11, no. 06, pp. 170 –177, 2003.

[12] B. Tessem, "Approximations for efficient computation in the theory of evidence," *Artificial Intelligence*, vol. 61, pp. 315 – 329, 1993.

[13] F. Cuzzolin, "Two new bayesian approximations of belief functions based on convex geometry," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. Volume 37, no. 4, pp. 993–1008, 2007.

[14] ——, "Lp consonant approximation of belief functions in the mass space," in *Proceedings of the 7th International Symposium on Imprecise Probability: Theories and Applications*, Innsbruck, Austria., 2011, pp. 1 – 10.

[15] ——, "On consistent approximations of belief functions in the mass space," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, ser. Lecture Notes in Computer Science, W. Liu, Ed. Springer Berlin Heidelberg, 2011, vol. 6717, pp. 287–298.

[16] A.-L. Jousselme and P. Maupin, "Distances in evidence theory : Comprehensive survey and generalizations," *International Journal of Approximate Reasoning*, vol. 53, pp. 118 – 145, 2012.

[17] M. Loudahi, J. Klein, J.-M. Vannobel, and O. Colot, "New distances between bodies of evidence based on dempsterian specialization matrices and their consistency with the conjunctive combination rule," *International Journal of Approximate Reasoning*, vol. 55, no. 5, pp. 1093 – 1112, 2014.

[18] ——, "Evidential matrix metrics as distances between meta-data dependent bodies of evidence," *IEEE Transactions on Cybernetics*, vol. accepted for publication, pp. 1–14, 2015.

[19] F. Cuzzolin, *Visions of a generalized probability theory*. Lambert Academic Publishing, 2014.

[20] G. J. Klir, *Uncertainty and Information: Foundations of Generalized Information Theory*. Wiley-IEEE Press, december 2005.

[21] D. Dubois and H. Prade, *Fuzzy Information and Decision Processes*. Amsterdam: North-Holland Publishing Company, 1982, ch. On several representations of an uncertain body of evidence, pp. 161–181.

[22] F. Cuzzolin, "A geometric approach to the theory of evidence," *IEEE Transactions on Systems, Man and Cybernetics - Part C: Application And Reviews*, vol. 38, no. 4, pp. 522 – 534, 2008.

[23] T. Denœux, "Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence," *Artificial Intelligence*, vol. 172, pp. 234 – 264, 2008.

[24] D. Dubois and H. Prade., "A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets." *International Journal of General Systems*, vol. 12, no. 3, pp. 193–226, 1986.

[25] F. Klawonn and P. Smets, "The dynamic of belief in the transferable belief model and specialization-generalization matrices," in *Proceedings of the eighth International Conference on Uncertainty in Artificial Intelligence*, D. Dubois, M. P. Wellman, B. d'Ambrosio, and P. Smets, Eds., vol. 92. San Mateo, CA. USA: Morgan Kaufmann Publishers Inc., 1992, pp. 130 –137.

[26] P. Smets, "The application of the matrix calculus to belief functions," *International Journal of Approximate Reasoning*, vol. 31, pp. 1–30, 2002.

[27] A.-L. Jousselme, D. Grenier, and E. Bossé, "A new distance between two bodies of evidence," *Information Fusion*, vol. 2, pp. 91–101, 2001.

[28] J. Diaz, M. Rifqi, and B. Bouchon-Meunier, "A similarity measure between basic belief assignments," in *Proceedings of the 9th International Conference on Information Fusion (FUSION 2006)*, Florence, Italy, 2006, pp. 1–6.

[29] S. Destercke and D. Dubois, "Idempotent conjunctive combination of belief functions: Extending the minimum rule of possibility theory," *Inf. Sci.*, vol. 181, no. 18, pp. 3925–3945, 2011.

[30] D. Dubois and H. Prade, "Consonant approximations of belief functions," *International Journal of Approximate Reasoning*, vol. 4, no. 56, pp. 419 – 449, 1990.

[31] F. Cuzzolin, "Geometric conditioning of belief functions," in *Proceedings of BELIEF 2010, International workshop on the theory of belief functions*, Brest, France, 2010, pp. 1–6.

[32] ——, "The geometry of consonant belief functions: Simplicial complexes of necessity measures," *Fuzzy Sets and Systems*, vol. 161, no. 10, pp. 1459 – 1479, 2010.

[33] ——, "Geometry of Dempster's rule of combination," *IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics*, vol. 34, no. 2, pp. 961 – 977, 2004.

[34] J. Abellan and S. Moral, "A non-specificity measure for convex sets of probability distributions," *International journal of uncertainty, fuzziness and knowledge-based systems*, vol. 8, no. 03, pp. 357–367, 2000.

[35] J. Abellán and S. Moral, "Difference of entropies as a non-specificity function on credal sets," *International journal of general systems*, vol. 34, no. 3, pp. 201–214, 2005.

[36] S. Destercke and T. Burger, "Toward an axiomatic definition of conflict between belief functions," *Cybernetics, IEEE Transactions on*, vol. 43, no. 2, pp. 585–596, April 2013.

[37] T. Burger, "Geometric interpretations of conflict: A viewpoint," in *Belief Functions: Theory and Applications*, ser. Lecture Notes in Computer Science, F. Cuzzolin, Ed. Springer International Publishing, 2014, vol. 8764, pp. 412–421.

[38] T. Denœux, "Extending stochastic ordering to belief functions on the real line," *Information Sciences*, vol. 179, no. 9, pp. 1362–1376, 2009.

[39] D. Dubois, H. Prade, and P. Smets, "New semantics for quantitative possibility theory," in *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, ser. Lecture Notes in Computer Science, S. Benferhat and P. Besnard, Eds. Springer Berlin Heidelberg, 2001, vol. 2143, pp. 410–421.

[40] ——, "A definition of subjective possibility," *International Journal of Approximate Reasoning*, vol. 48, no. 2, pp. 352 – 364, 2008, in Memory of Philippe Smets (19382005).

University of Lille, France

Technologic University of Compiegne,France

E-mail address: {john.klein,olivier.colot}@univ-lille1.fr   sebastien.destercke@hds.utc.fr