International Journal of Approximate Reasoning 53 (2012) 118–145

Contents lists available at SciVerse ScienceDirect

International Journal of Approximate Reasoning

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

Distances in evidence theory: Comprehensive survey and generalizations ☆

Anne-Laure Jousselme *, Patrick Maupin

Defence R&D Canada – Valcartier, Decision Support Systems for Command and Control (DSS-C2) Section, 2459 Pie XI North, Quebec, QC, Canada G3J 1X5

ARTICLE INFO

Article history:
Available online 31 August 2011

Keywords:
Dissimilarity
Distance
Metric
Inner product
Belief functions
Additive tree

ABSTRACT

The purpose of the present work is to survey the dissimilarity measures defined so far in the mathematical framework of evidence theory, and to propose a classification of these measures based on their formal properties. This research is motivated by the fact that while dissimilarity measures have been widely studied and surveyed in the fields of probability theory and fuzzy set theory, no comprehensive survey is yet available for evidence theory. The main results presented herein include a synthesis of the properties of the measures defined so far in the scientific literature; the generalizations proposed naturally lead to additions to the body of the previously known measures, leading to the definition of numerous new measures. Building on this analysis, we have highlighted the fact that Dempster's conflict cannot be considered as a genuine dissimilarity measure between two belief functions and have proposed an alternative based on a cosine function. Other original results include the justification of the use of two-dimensional indexes as (cosine; distance) couples and a general formulation for this class of new indexes. We base our exposition on a geometrical interpretation of evidence theory and show that most of the dissimilarity measures so far published are based on inner products, in some cases degenerated. Experimental results based on Monte Carlo simulations illustrate interesting relationships between existing measures.

1. Introduction

This paper presents, together with a synthesis of their main properties, a survey of the main dissimilarity measures defined so far using the mathematical framework of evidence theory. This work leads to the definition of a reduced set of general formulations, allowing the authors to categorize existing dissimilarity measures and in addition, to define several new measures. Numerous technical observations are made about the surveyed measures and Monte Carlo simulations are used to illustrate important differences in practical behaviors. An analysis of these experimentally outlined differences is made and parallels can be drawn with the previously obtained categorization based on theoretical properties.

It shall be noted that we use the term "distance" herein to designate the intuitive notion relative to a somewhat quantified difference among a set of objects, even though in the technical literature the term is often used to designate an actual measure having precise properties. We prefer here to use the technical terms of dissimilarity and similarity, although in the current section the distinction will not be clear cut as we will respect as much as possible the terminology used in the works referred to.

The vast body of literature on the evaluation of distances between probability distributions is of course a great source of inspiration for the definition of dissimilarity measures in evidence theory. Among the reference papers also aimed at studying dissimilarities in the broad sense, we can cite the works of Basseville [2] and a recent paper of Cha [11] who proposes a classification of the main distances in probability theory as well as a comparison method for the distances.

☆ A preliminary version of this paper has been presented at WBF'10 [36].

* Corresponding author.
E-mail address: Anne-Laure.Jousselme@drdc-rddc.gc.ca (A.-L. Jousselme).

0888-613X/$ - see front matter Crown copyright © 2011 Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.ijar.2011.07.006

In fuzzy set theory, Bloch proposed a detailed survey of distances between fuzzy sets in [8] where fuzzy distances are reviewed and classified with respect to the needs in image processing applications. In the framework of imprecise probabilities, Abellán and Gómez [1] defined three measures for comparison of credal sets [1] : an inconsistency measure, an inclusion index and an informative distance. In [10], de Campos et al. proposed a method for building distances between fuzzy measures based on associated probability distributions. In possibility theory, Jenhani et al. defined in [34] a distance between possibility distributions together with some required properties.

In recent years, many works on measuring the distance between belief functions have emerged. For a long time, Demspter's conflict factor has been the only way to quantify the interaction between belief functions (see for example [54,23]). However, this factor may not be appropriate to quantify the dissimilarity between two belief functions as the conflict between two identical belief functions may not equal to 0, a result somewhat counterintuitive. Several approaches have been proposed for the definition of distances in evidence theory. In [56,46], Perry and Stephanou extended the Kullback-Liebler divergence for probability distributions, Blackman and Popoli [7] and Ristic and Smets [50] defined a distance based on Dempster's conflict factor. Other authors proposed geometrical (Euclidean) distances: Fixsen and Mahler [26] defined a classification miss-distance, Jousselme et al. [35] proposed a geometric distance accounting for the similarity between focal sets, Cuzzolin [14] defined an Euclidean measure between belief values and extended it to L_p Minkowski measures in [16], Wen proposed to quantify the similarity as the cosine measure of the angle between two mass vectors [59].

In the technical literature, two main aims may be identified regarding the practical use of distances between belief functions: (1) for algorithm evaluation or optimization, for example in classification algorithms [26,35,19], or in belief functions approximation algorithms [57,3,14,20], or for combination rule parameter estimation [21,48,61], (2) as a definition of agreement between sources of information, for example in clustering techniques [4,60,5,52], or as a basis for discounting factors [43,29,12,31,38,45]. In algorithm evaluation or optimization, the distance is computed with respect to a reference belief function Bel^r or to an entire reference space, as in works on approximation algorithms where distances are measured with respect to the subspace of Bayesian or consonant belief functions ([13] for instance), whereas in the definition of agreement between sources of information no such reference exists. Depending on the application, some formal properties are required while some others are superfluous. Our position is that none of the distance measures can be said to be superior to the others in the absolute and that the choice of such a measure should always be guided by practical considerations relative to a specific application.

In Section 2, we review important basic notions of evidence theory, including notational conventions that will ease the subsequent analytic exposition, together with an emphasis on the geometrical interpretation in Section 2.2. The properties of similarity and dissimilarity measures are detailed in Section 2.3. A categorization of the distances is proposed in Section 3, based on the definition a set of different inner-products between belief functions. It is shown that most of the existing measures can fit in this general formulation. We also mention other works of interest. In Section 4, we discuss some outcomes of the present survey: Section 4.1 summarizes the existing measures and proposes new measures, Section 4.2 discusses the metric and structural properties, the normalization factors are presented in Section 4.3, Bayesian belief functions are discussed in 4.4, Section 4.5 hightlights two kinds of measures that are metric distances and angles, an alternative to Dempster's rule is proposed in Section 4.6, different encoding of belief functions are proposed in Section 4.7 and some comments about extensions to unnormalized and fuzzy belief functions are provided in Section 4.8. An experimental comparison of distance measures is proposed in Section 5 firstly based on a toy example (Section 5.1) and on a more semantic approach using additive trees (Section 5.2). Section 6 concludes on future works that will be developed in upcoming publications.

2. Background

The background material presented in this section deals with the following four main points: (1) the geometrical interpretation of belief functions, that will be used in this paper to ease the exposition, (2) definitions on similarity, dissimilarity and metric measures, (3) basic notions on inner products and distances, and (4) the structural properties of belief functions.

2.1. Basics on evidence theory

Let X be a frame of discernment containing N distinct objects x_i, $i = 1, \dots, N$. We denote by x an element of X. The power set of X, denoted by 2^X, is the set of the 2^N subsets of X. A Basic Probability Assignment (BPA) m is a mapping from 2^X to $[0, 1]$ satisfying the two following conditions:

$$\sum_{A\subseteq X} m(A) = 1 \quad \text{and} \quad m(\varnothing) = 0 \tag{1}$$

A subset A of X is called a *focal element* if $m(A) > 0$ and we denote by $\mathcal{F}$ the *set* of all the focal elements, *i.e.* $\mathcal{F} = \{A \subseteq X | m(A) > 0\}$. Three one-to-one mappings can be defined from m, namely the *belief*, *plausibility* and *commonality* respectively for all $A \subseteq X$:

[1] A credal set K is a closed and convex set of probability distributions. The credal set associated with a belief function Bel defined on a frame of discernment X is $K = \{p|\mathrm{Bel}(A) \leq p(A), \forall A \subseteq X\}$.

$$\mathrm{Bel}(A) = \sum_{B \subseteq A} m(B), \quad \mathrm{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B), \quad q(A) = \sum_{A \subseteq B} m(B) \tag{2}$$

The *pignistic probability* [55] is defined for all A of X by:

$$\mathrm{Bet\,P}(A) = \sum_{B \subseteq X} m(B) \frac{|A \cap B|}{|B|} \tag{3}$$

where $|A|$ is the cardinality of set A. In particular, if A is a singleton $\{x\}$, we have $\mathrm{Bet\,P}(\{x\}) = \sum_{x \in B} \frac{m(B)}{|B|}$.

Let us introduce the following indexes between two subsets of X:

$$\text{Inclusion index}: \ \mathrm{Inc}(A, B) = 1 \text{ if } A \subseteq B \text{ and 0 otherwise} \tag{4}$$

$$\text{Intersection index}: \ \mathrm{Int}(A, B) = 1 \text{ if } A \cap B \neq \emptyset \text{ and 0 otherwise} \tag{5}$$

$$\text{Pignistic index}: \ \mathrm{Bet}(A, B) = \frac{|A \cap B|}{|B|} \tag{6}$$

The Inc index corresponds to **QfrM** in [53]. We note that Int is symmetric while Inc and Bet are not. The dual index of Int is $1 - \mathrm{Int}$ which is such that $1 - \mathrm{Int}(A, B) = 1$ iff $A \cap B = \emptyset$ and 0 otherwise. Introducing these indexes allows alternative notations for Eqs. (2) and (3):

$$\mathrm{Bel}(A) = \sum_{B \subseteq X} m(B)\mathrm{Inc}(B, A) \tag{7}$$

$$\mathrm{Pl}(A) = \sum_{B \subseteq X} m(B)\mathrm{Int}(A, B) \tag{8}$$

$$q(A) = \sum_{B \subseteq X} m(B)\mathrm{Inc}(A, B) \tag{9}$$

$$\mathrm{Bet\,P}(A) = \sum_{B \subseteq X} m(B)\mathrm{Bet}(A, B) \tag{10}$$

2.2. A geometrical interpretation of evidence theory

The geometrical interpretation of evidence theory can be traced back to the work of Ronald Mahler in 1996 [42], where the author sets the bases with a random sets interpretation of belief functions.[2] This interpretation has also been used in [35] to define a distance between two belief functions, and further developed by Cuzzolin in [14,15] for instance.

Let $\mathcal{E}_X$ be the 2^N-dimensional Cartesian space spanned by the set of vectors $\{\mathbf{e}_A, A \subseteq X\}$. Any vector $\mathbf{v}$ of $\mathcal{E}_X$ can be then written as $\mathbf{v} = \sum_{A \subseteq X} \alpha_A \mathbf{e}_A$, where $\alpha_A \in \mathbf{R}$ is the coordinate of $\mathbf{v}$ along the dimension $\mathbf{e}_A$.

A BPA is a vector[3] $\mathbf{m}$ of $\mathcal{E}_X$ satisfying the properties (1), *i.e.* $\sum_{A \subseteq X} \alpha_A = 1$, $\alpha_\emptyset = 0$, with $\alpha_A \geq 0$ together with $\alpha_A = m(A)$. A belief function Bel is then represented equivalently by a vector $\mathbf{Bel} = \sum_{A \subseteq X} \mathrm{Bel}(A)\mathbf{e}_A$, with its belief values $\mathrm{Bel}(A)$ as coordinates of **Bel**. Equivalent representations hold for **Pl** and **q**.

Using a vector-matrix notation as proposed in [53] is natural and makes the exposition easier. Let us now denote by **Inc**, **Int** and **Bet** as being the matrices whose elements are defined by Eqs. (4)–(6), and let $\mathbf{Inc}'$ denote the transpose matrix of **Inc**. We can then rewrite Eqs. (7)–(10) as the following products:

$$\mathbf{Bel} = \mathbf{Inc}' \cdot \mathbf{m}, \quad \mathbf{Pl} = \mathbf{Int} \cdot \mathbf{m}, \quad \mathbf{q} = \mathbf{Inc} \cdot \mathbf{m} \quad \mathbf{BetP} = \mathbf{Bet} \cdot \mathbf{m} \tag{11}$$

For example, for $N = 2$ we have:

$$\mathbf{Inc} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Int} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad \mathbf{Bet} = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \\ 1 & 1 & 1 \end{pmatrix} \tag{12}$$

where the focal elements have been ordered as $\{x_1, x_2, (x_1, x_2)\}$ in rows and in columns. The matrix **Inc** is a upper triangular matrix with positive values on the diagonal. The matrix **Bet** is of rank N (instead of $2^N - 1$ if it had be of full rank) because $\mathrm{Bet\,P}(A) = \sum_{x \in A} \mathrm{Bet\,P}(\{x\})$. It defines thus a projection over $\mathcal{E}_x$, the N-dimensional subspace of $\mathcal{E}_X$ spanned by the singleton vectors $\mathbf{e}_x$, *i.e.* the space of probability distributions. We will denote in the following by $\mathbf{Bet}_x$ the rectangular matrix with N rows and $2^N - 1$ columns. Then, $\mathbf{BetP} = \mathbf{Bet}_x \cdot \mathbf{m}$ is thus the pignistic Bayesian approximation (or pignistic transformation)

[2] Note that the geometrical interpretation referred here should not be confounded with the work of Kendall and Mathéron [37,44] on random sets, as the latter work provides a geometrical setting for random sets in $\mathbf{R}^d$ and not to probability distributions explicitly.

[3] As a convention, a vector $\mathbf{v}$ is a column vector, and its transpose $\mathbf{v}'$ a row vector.

Table 1
Axioms for metrics. The properties hold for all $(y, z, t) \in \mathcal{S}^3$.

		Metric	Semi-metric	Quasi-metric	Pseudo-metric	Semi-pseudo-metric	Pre-metric
(d1) Nonnegativity	$d(y,z) \geq 0$	×	×	×	×	×	×
(d2) Symmetry	$d(y,z) = d(z,y)$	×	×		×	×	
(d3) Definiteness	$d(y,z) = 0 \Leftrightarrow y = z$	×	×	×			
(d3)' Reflexivity	$d(y,y) = 0$	×	×	×	×	×	×
(d3)" Separability	$d(y,z) = 0 \Rightarrow y = z$	×	×	×			
(d4) Triangle inequ.	$d(y,z) \leq d(y,t) + d(t,z)$	×		×	×		

of Bel, and **BetP** is a vector of size $N \times 1$. In general, we will denote by $\mathbf{W}_x$ the $N \times 2^N - 1$ rectangular matrix corresponding to the restriction of the square matrix **W** to the N singletons ($\mathbf{Bet}_x$, $\mathbf{Inc}_x$, $\mathbf{Int}_x$, ...). Note that **Bet** is denoted by **BetPfrM** and $\mathbf{Bet}_x$ by **betPfrM** in [53].

2.3. Similarities and dissimilarities

In this paper, the term "distance" is used to denote a general and intuitive notion of distance quantifying how much two objects are different. This intuitive notion encompasses all the other more formal terms used in the paper. In particular, we distinguish formal notions defined by axiomatic properties of Table 1 such as (full) metric,[4] pseudometric, semimetric. All other distances satisfying less axioms than a premetric will be called nonmetric.

Let y be an element of a given space $\mathcal{S}$. In this paper, $\mathcal{S}$ will represent either $\mathcal{E}_X$, the space of belief functions in which case $y \equiv m$, or 2^X, the set of focal elements in which case $y \equiv A$.

A function $d : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ is a (*full*) *metric* if and only if d satisfies the following properties for all $(y, z, t) \in \mathcal{S}^3$:

(d1) Nonnegativity: $d(y,z) \geq 0$.
(d2) Symmetry: $d(y,z) = d(z,y)$.
(d3) Definiteness: $d(y,z) = 0 \Leftrightarrow y = z$.
(d4) Triangle inequality: $d(y,z) \leq d(y,t) + d(z,t), \ \forall t$.

Property (d3) can be split into (d3)' and (d3)", the properties of:

(d3)' Reflexivity[5]: $d(y,y) = 0$.

(d3)" Separability: $d(y,z) = 0 \Rightarrow y = z$,

(d1) together with (d3) define positive definiteness.

If d satisfies only some subsets of the set of axioms above, then different degenerate forms of metrics are defined, as summarized in Table 1.

The weakest kind of metric is a *premetric* which is only nonnegative (d1) and reflexive (d3)'. A *quasimetric* is all excepted that it is not symmetric. A *pseudometric* is almost a (full) metric excepted that it does not satisfy the separability property (d3)", and consequently, the definiteness property (d3). The separability property guaranties that a null distance is obtained only between an object and itself. If not satisfied, two distinct objects may have a null distance. A *semimetric* satisfies all the properties excepted the triangular inequality (d4). And a (full) metric satisfies all axioms from (d1) to (d4). Some combinations of the prefixes semi, pre, quasi, pseudo may be combined leading to other kinds of degenerate metrics. Of particular interest in this paper, a *semipseudometric* is a distance satisfying all five axioms except (d4) (semi) and (d3)" (pseudo). A distance which does not minimally satisfy the two axioms of nonnegativity and reflexivity will be called *nonmetric*.

A function $s : \mathcal{S} \times \mathcal{S} \to \mathbb{R}$ is a *similarity* if and only if s satisfies the following properties for all $(y, z) \in \mathcal{S}^2$:

(s1) Symmetry[6]: $s(y,z) = s(z,y)$,
(s2) $s(y,y) \geq s(y,z)$, for all $y \neq z$.

Furthermore if s satisfies:

(s3) Normality: $s(y,y) = 1$,

then, s is *normed*. A *dissimilarity* minimally satisfies the axioms (d1), (d2) and (d3)'. Several techniques exist that allow the definition of dissimilarities from similarities (see [30]), like the simple relation $d = 1 - s$. For instance, the cosine measure which is known to be a similarity measure can be transformed into a dissimilarity measure by defining $\cos^d(\theta) = 1 - \cos(\theta)$.

[4] Although being redundant, the term "full metric" will be sometimes used in this paper to distinguish it from a pseudometric.
[5] The reflexivity axiom is also called *identity* axiom. Also, note that a weakest form of the reflexivity axiom can be defined as $d(y,y) = \alpha$ where α is a constant.
[6] Note that in some cases, like in directional statements involving an object and a referent (*e.g.*, "*A* is like *B*"), the symmetry property may be considered too strong for similarity measures [58].

2.4. Inner products and distances

An inner product $\otimes$ over a linear space $\mathcal{V}$ is a mapping

$$\begin{aligned}\otimes :\ &\mathcal{V}\times\mathcal{V}\longrightarrow \mathbf{R}\\ &(\mathbf{v}_1,\mathbf{v}_2)\longmapsto \otimes(\mathbf{v}_1,\mathbf{v}_2)=\alpha\end{aligned}$$

which must satisfy the 3 following axioms for all vectors $\mathbf{v}_1,\mathbf{v}_2,\mathbf{v}_3$ of $\mathcal{V}$ and all scalars $a, b\in\mathbf{R}$:

(ip1) Symmetry: $\otimes(\mathbf{v}_1,\mathbf{v}_2)=\otimes(\mathbf{v}_2,\mathbf{v}_1)$,
(ip2) Linearity in the first argument: $\otimes(a\mathbf{v}_1+b\mathbf{v}_2,\mathbf{v}_3)=a\otimes(\mathbf{v}_1,\mathbf{v}_3)+b\otimes(\mathbf{v}_2,\mathbf{v}_3)$,
(ip3) Positive-definiteness: $\otimes(\mathbf{v},\mathbf{v})\geq 0$ with equality only for $\mathbf{v}=\mathbf{0}$.

A linear space $\mathcal{V}$ endowed with an inner product $\otimes$ is called an *inner product space*. This is true for $\mathcal{E}_X$ in particular.

A general representation for an inner product is:

$$\otimes_W(\mathbf{v}_1,\mathbf{v}_2)=\mathbf{v}_1'\mathbf{W}\mathbf{v}_2 \tag{13}$$

where $\mathbf{W}$ is a matrix of weights (weighting matrix) required to be symmetric and positive definite.[7] The angle between $\mathbf{v}_1$ and $\mathbf{v}_2$ is given by:

$$\theta=\arccos\left(\frac{\otimes_W(\mathbf{v}_1,\mathbf{v}_2)}{\|\mathbf{v}_1\|_W.\|\mathbf{v}_2\|_W}\right) \tag{14}$$

where the *norm* of a vector $\mathbf{v}$ is defined as $\|\mathbf{v}\|_W=\sqrt{\otimes(\mathbf{v},\mathbf{v})}$ and represents the length of $\mathbf{v}$. The norm can be used to define a *distance function* on $\mathcal{E}_X$ by:

$$d(\mathbf{v}_1,\mathbf{v}_2)=\|\mathbf{v}_1-\mathbf{v}_2\|_W=\sqrt{(\mathbf{v}_1-\mathbf{v}_2)'\mathbf{W}(\mathbf{v}_1-\mathbf{v}_2)} \tag{15}$$

An inner product is *degenerate* if it satisfies all the properties except the separability property, *i.e.* if $\|\mathbf{v}\|_W=0$ does not imply $\mathbf{v}=0$. In this case, $\mathbf{W}$ is only positive semidefinite. The induced norm is then a *pseudonorm* and the induced distance, a *pseudometric* (see Table 1).

If $\mathbf{W}$ is square, symmetric and positive definite, then $\mathbf{W}$ can be uniquely factorized as $\mathbf{W}=\mathbf{U}'\mathbf{U}$, where $\mathbf{U}$ is upper triangular with positive diagonal entries (Cholesky decomposition). Note that this result also holds whenever $\mathbf{W}$ is positive semidefinite. Then, we can write (13) and (15), respectively as:

$$\otimes(\mathbf{v}_1,\mathbf{v}_2)=(\mathbf{U}\mathbf{v}_1)'(\mathbf{U}\mathbf{v}_2) \tag{16}$$

and

$$d(\mathbf{v}_1,\mathbf{v}_2)=\sqrt{(\mathbf{U}(\mathbf{v}_1-\mathbf{v}_2))'\,\mathbf{U}(\mathbf{v}_1-\mathbf{v}_2)} \tag{17}$$

2.5. Structural property of belief functions

Besides the axioms introduced in the previous section, a distance between belief functions should satisfy other properties very specific to the nature of belief functions. Although the development of such properties is out of the scope of the present paper, we will however consider the following properties:

(sp1) Strong structural property (*interaction between focal elements*):
A distance measure d between two belief functions Bel_1 and Bel_2 is said to be *strongly structural* if its definition accounts for the interaction between the focal elements of Bel_1 and Bel_2, *i.e.* if $s(\mathbf{e}_A,\mathbf{e}_B)$, where s quantifies the interaction between the basis vectors, plays a role in the definition of d.

(sp2) Weak structural property (*cardinality of focal elements*):
A distance measure d between two belief functions Bel_1 and Bel_2 is said to be *weakly structural* if its definition accounts for the cardinality of the focal elements of Bel_1 and Bel_2, *i.e.* if $|A|$ plays a role in the definition of d.

(sp3) Structural dissimilarity (interaction between sets of focal elements):
A distance measure d between two belief functions Bel_1 and Bel_2 is said to satisfy the *structurally dissimilarity* property if its definition accounts for the interaction between the sets $\mathcal{F}_1$ and $\mathcal{F}_2$ of focal elements of Bel_1 and Bel_2, *i.e.* if $d_{fe}(\mathcal{F}_1,\mathcal{F}_2)$, where d_{fe} quantifies the interaction between the sets of focal elements.

[7] A matrix is positive definite iff all its eigenvalues are strictly positive ($\lambda_i>0$), and it is positive semidefinite iff its eigenvalues are positive ($\lambda_i\geq 0$).

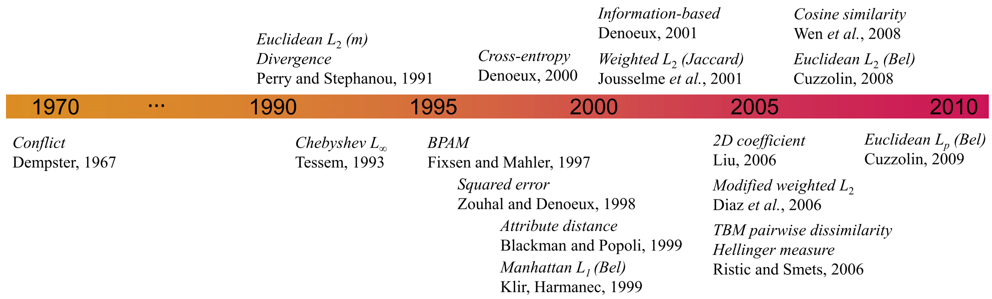


Fig. 1. Historical contributions to the measurement of distance between two belief functions.

Indeed, compared to traditional spaces, the basis vectors $\mathbf{e}_A$, $A \subseteq X$, of the space $\mathcal{E}_X$ of belief functions can be linked through some similarity or dissimilarity measures, making this space in some sense "curved". For instance, the basis vectors $\mathbf{e}_A$ and $\mathbf{e}_B$ where $A = \{x_1, x_2, x_3\}$ and $B = \{x_1, x_2\}$ are more similar than the basis vectors $\mathbf{e}_A$ and $\mathbf{e}_C$ where $C = \{x_4, x_5\}$. This structural property is an interesting property to be satisfied by a distance measure between belief functions and will be considered in the upcoming analysis as the axiomatic metric properties will be.

2.6. Notations

In the remaining of this paper, we will denote by d a dissimilarity measure, by s a similarity measure, by $\otimes$ an inner product, and by cos a cosine measure. Moreover, a subscript W will be added to the previous ones to specify the weighting matrix. A superscript s or d will be added to $\otimes$ to specify if $\mathbf{W}$ defines a similarity or dissimilarity respectively. A superscript (p) will also be added to d to denote the Minskowski family, $p \in \{1, 2, \ldots, \infty\}$.

3. Survey of the main distances in evidence theory and classification

Since the introduction of Dempster's conflict measure [17], about 20 distance measures between belief functions have been defined in the technical literature. Fig. 1 summarizes several contributions that will be reviewed in this paper. We hope this survey is exhaustive and apologize for any forgotten contribution.

Most of the distance measures defined so far in the evidence theory framework are derived from inner products, either directly (see Section 3.3), or through metrics. The main family of metrics considered is the Minkowski family (Section 3.2), denoted as L_p, whose most famous representative is the Euclidean metric family L_2 (see Section 3.2.2). L_1 metrics (Manhattan) and L_∞ metrics (Chebyshev norm, also called infinity, uniform or supremum norm) will be reviewed in Sections 3.2.1 and 3.2.3 respectively. The Fidelity family of distances will be described in Section 3.4 and information-based distances will be presented in Section 3.5. This categorization leads naturally to generalizations that will be made explicit in the following, and that will help us obtain the definition of more than 40 new measures of distances between belief functions in Section 4.1.

As an introduction to this section, we present two composite measures that cannot be formally classified into the metric family even if some of their individual components could be to some extent.

3.1. Composite distances

The first work addressing specifically the problem of quantifying the distance between two belief functions is that of Perry and Stephanou who have proposed in [46], based on [56], a measure of divergence between two belief functions to be used by a classifier. The authors argue that an evaluation of distance should measure "the difference between the amount of information available when they are considered separately and when they are combined". They thus proposed what they have called "an extension of the symmetric version of Kullback-Liebler divergence"[8] for probability distributions based on the fact that the updating rule is Dempster's combination rule rather than Bayes' rule:

$$d_{PS}(m_1, m_2) = \underbrace{|\mathcal{F}_1 \cup \mathcal{F}_2|\left(1 - \frac{|\mathcal{F}_1 \cap \mathcal{F}_2|}{|\mathcal{F}_1 \cup \mathcal{F}_2|}\right)}_{d_{PS(1)}} + \underbrace{(\mathbf{m}_{12} - \mathbf{m}_1)'(\mathbf{m}_{12} - \mathbf{m}_2)}_{d_{PS(2)}} \tag{18}$$

where $\mathcal{F}_i$ is the set of focal elements of m_i and $\mathbf{m}_{12}$ is the BPA obtained by combining $\mathbf{m}_1$ and $\mathbf{m}_2$ with Dempster's rule. The resulting expression has two components: (1) a measure of *structural dissimilarity* (dissimilarity between sets of focal elements), $d_{PS(1)}$ and (2) a measure of *information change* relatively to the orthogonal sum, $d_{PS(2)}$. The member $d_{PS(1)}$ quantifies

[8] The reader is referred to the original paper [56] for an argumentation of the authors in favour of this formulation.

how close the two sets of focal elements are from each other: If Bel_1 and Bel_2 have the same focal elements ($\mathcal{F}_1 = \mathcal{F}_2$), then $d_{PS(1)} = 0$ meaning that Bel_1 and Bel_2 are *structurally identical*. Referring to the discussion in Section 4.5 about the angles and distances, two belief functions with the same set of focal elements ($d_{PS(1)} = 0$) are found to be collinear.

The underlying intuition in Perry and Stephanou's divergence is that two aspects must be considered, namely the interaction between focal elements (the so-called structural property in the remainder of the present article) and the difference in mass values. The distance d_{PS} can be analyzed component by component: the first component $d_{PS(1)}$ satisfies the structural dissimilarity property (sp3) and the basic axioms (d1) and (d2). The second component $d_{PS(2)}$ satisfies the strong structural property (sp1) due to the use of Dempster's combination and the basic axiom (d2) but fails to satisfy the non-negativity axiom (d1). The global distance d_{PS} has thus at least four shortcomings: (1) the range of $d_{PS(1)}$ is much higher than that of $d_{PS(2)}$, which puts a (too) large emphasis on the structural property (larger if $|X|$ is high), (2) d_{PS} is a nonmetric measure as it does not satisfy the (d1), (d3)' ($d_{PS(2)}(m, m) \neq 0$), (d3)" and (d4) axioms, (3) $d_{PS(2)}$ (thus d_{PS}) is not defined if the (Dempster's) conflict between m_1 and m_2 is 1 (as is m_{12}), (4) d_{PS} is undefined for $d_{PS}(\mathbf{e}_A, \mathbf{e}_B)$ such that $A \cap B = \emptyset$.

In [7], Blackman and Popoli defined what they called an "attribute distance" to be used in association algorithms:

$$d_{BP}(m_1, m_2) = \underbrace{-2 \log \left[\frac{1 - \otimes^d_{Int}(m_1, m_2)}{1 - \max_{i=1,2} \otimes^d_{Int}(m_i, m_i)} \right]}_{d_{BP(1)}} + \underbrace{(\mathbf{m}_1 + \mathbf{m}_2)' \mathbf{g}_A - \mathbf{m}_1' \mathbf{G} \mathbf{m}_2}_{d_{BP(2)}} \tag{19}$$

where $\otimes^d_{Int}(m_1, m_2)$ is Dempster's conflict introduced in (43), $\mathbf{g}_A$ is a vector whose elements are $\frac{|A|-1}{|X|-1}$, and $\mathbf{G} = \mathbf{g}_A \mathbf{g}_A'$ is a matrix whose elements are $G(A, B) = \frac{(|A|-1)(|B|-1)}{(|X|-1)^2}$, $A, B \subseteq X$, where $\mathbf{m}'.\mathbf{g}_A$ is the *partial ignorance* introduced in [56].

The first component of (19) (denoted by $d_{BP(1)}$ hereafter) has been called "attribute distance" by the authors while the second member (denoted by $d_{BP(2)}$ hereafter) has been called "ignorance distance". The quantity $d_{BP(1)}$ remains undefined whenever $\otimes^d_{Int}(m_1, m_2) = 1$ (total conflict) and is equal to zero whenever $\otimes^d_{Int}(m_1, m_2) = 0$ (null conflict). However, a null conflict (in Dempster's sense) between m_1 and m_2 does not imply that $m_1 = m_2$. The second term $d_{BP(2)}$ serves as a penalty factor aiming at the discrimination between cases of perfect match and those depicting an ignorance situation. Summing up these two components leads to a non positive measure, and thus d_{BP} is a nonmetric distance (axiom (d1) is not satisfied). In fact, (d2) the symmetry axiom is the only axiom satisfied by d_{BP}.

Composite distances lead to deceiving results in terms of metric properties as they only satisfy the symmetry axiom (d2). However, the individual components involved, *i.e.* $d_{PS(1)}, d_{PS(2)}, d_{BP(1)}, d_{BP(2)}$, are interesting since each highlights some requirements in particular, a distance measure between belief functions should account in an *aggregate manner* for both the *structural properties* (difference in the $\mathcal{F}_i$s) and the *mass dissimilarity* (difference in the $m_i(A_j)$). Most of the distances presented in the following aim at addressing these requirements. We have been inspired by [11] for the structure and terminology used in the upcoming subsections.

3.2. Minkowski family

The Minkowski family of distances between two belief functions can be written under the following general form:

$$d^{(p)}_W(m_1, m_2) = \left(\left[(\mathbf{U}\mathbf{m}_1 - \mathbf{U}\mathbf{m}_2)^{\frac{p}{2}} \right]' \left[(\mathbf{U}\mathbf{m}_1 - \mathbf{U}\mathbf{m}_2)^{\frac{p}{2}} \right] \right)^{\frac{1}{p}} \tag{20}$$

where $\mathbf{U}$ is the upper triangular matrix of the Cholesky decomposition of the matrix $\mathbf{W}$, that is $\mathbf{W} = \mathbf{U}'\mathbf{U}$, and p is an integer higher than 1. If $\mathbf{v}' = [v_1 \ldots v_N]$, then $\mathbf{v}^\alpha$ for $\alpha \in \mathbf{R}$ is the vector whose components are v_i^α. Typical cases of interest are obtained with $p = 1$, $p = 2$ and $p = \infty$ leading to the Manhattan (or city-block), Euclidean and Chebyshev distances respectively. Recently, Cuzzolin defined in [16] L_p measures between belief functions to address the problem of consistent approximations of belief functions:

$$d^{(p)}_{Inc}(m_1, m_2) = \left(\sum_{A \subseteq X} |\mathrm{Bel}_1(A) - \mathrm{Bel}_2(A)|^p \right)^{\frac{1}{p}} \tag{21}$$

This is just a special case of the general formulation (20) with $\mathbf{U} = \mathbf{Inc}'$.

3.2.1. Manhattan distances – L_1

The general L_1 measure for belief functions is obtained for $p = 1$ in (20):

$$d^{(1)}_W(m_1, m_2) = \left[(\mathbf{U}\mathbf{m}_1 - \mathbf{U}\mathbf{m}_2)^{\frac{1}{2}} \right]' \left[(\mathbf{U}\mathbf{m}_1 - \mathbf{U}\mathbf{m}_2)^{\frac{1}{2}} \right] \tag{22}$$

We obtain for $\mathbf{U} = \mathbf{Inc}'$:

$$d^{(1)}_{Inc}(m_1, m_2) = \sum_{A \subseteq X} |\mathrm{Bel}_1(A) - \mathrm{Bel}_2(A)| \tag{23}$$

This distance of type L_1 has been introduced by Klir and Harmanec in [40,32] and further used in [20] as an error measure for belief function approximations. Note however that the original formulation of (23) in [40] did not use the absolute value because the distance was defined for approximations of belief functions purposes and Bel_2 was a necessity measure consistent with Bel_1 thus such that $\mathrm{Bel}_1(A) \geq \mathrm{Bel}_2(A), \forall A \subseteq X$. We added the absolute value for the definition in the general case. In [14], Cuzzolin identified $d^{(1)}_{Inc}$ as inappropriate for the probabilistic approximation of a belief function as all Bayesian belief functions consistent with a given one Bel_0 have the same $d^{(1)}_{Inc}$ distance from Bel_0. In [20], Denœux introduced the counterpart of (23) for plausibility functions, obtained by letting $\mathbf{U} = \mathbf{Int}$:

$$d^{(1)}_{Int}(m_1, m_2) = \sum_{A \subseteq X} |\mathrm{Pl}_1(A) - \mathrm{Pl}_2(A)| \tag{24}$$

This distance measurement has been used in [20,22,16] for belief functions approximations. Also, in [45] a restricted version of (24) to singletons is used, thus with $\mathbf{U} = \mathbf{Int}_x$:

$$d^{(1)}_{Intx}(m_1, m_2) = \sum_{x \in X} |\mathrm{Pl}_1(x) - \mathrm{Pl}_2(x)| \tag{25}$$

As already noticed in [20], the duality of the Bel and Pl measures, *i.e.* $\mathrm{Pl}(A) = 1 - \mathrm{Bel}(\bar{A})$, makes the expressions (23) and (24) equal while it is no longer true under the open world assumption (unnormalized belief functions). Also, this property does not hold for (25) as $d^{(1)}_{Intx}(m_1, m_2) \neq d^{(1)}_{Incx}(m_1, m_2)$.

3.2.2. Euclidean distances – L_2

When $p = 2$, (20) becomes:

$$d^{(2)}_{W}(m_1, m_2) = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)'\mathbf{W}(\mathbf{m}_1 - \mathbf{m}_2)} \tag{26}$$

where $\mathbf{W} = \mathbf{U}'\mathbf{U}$ is a positive semidefinite matrix. Several L_2 distances (hence several inner products) can then be obtained by modifying the weighting matrix $\mathbf{W}$.

The simplest inner product in $\mathcal{E}_X$ is:

$$\otimes^s_I(m_1, m_2) = \mathbf{m}_1' \ \mathbf{I} \ \mathbf{m}_2 \tag{27}$$

where $\mathbf{I}$ is the identity matrix. The inner product $\otimes^s_I$ does not satisfy any structural property as it only accounts for the mass distribution over the focal elements and not for the interaction between the focal elements themselves. Like its inner product, the associated distance $d^{(2)}_I$ suffers from the same drawback. For example, the BPAs $m_1(\{x_1, x_2, x_3\}) = 0.8$, $m_1(\{x_2, x_3\}) = 0.2$ and $m_2(\{x_1, x_2, x_3, x_4\}) = 1$ are very far from each other according to $d^{(2)}_I$, while intuitively they are not. $d^{(2)}_I$ has been introduced first in [46] and used for instance in [19] as an optimization criterion in the training of a neural network, in [49] in a process of combination of pairwise classifiers or in [50] in an association algorithm.[9] However, one interesting property of this distance is that it is definite, and thus $d^{(2)}_I(m_1, m_2) = 0 \Leftrightarrow m_1 = m_2$. In fact $d^{(2)}_I$ satisfies all the axioms,(d1) to (d4), of a full metric.

In order to evaluate the performance of identification algorithms, Fixsen and Mahler proposed in [26] a "classification miss-distance metric" called BPAM, for Bayesian Percent Attribute Miss based on the following inner product:

$$\otimes^s_P(m_1, m_2) = \mathbf{m}_1' \ \mathbf{P} \ \mathbf{m}_2 \tag{28}$$

where $\mathbf{P}$ is the matrix whose elements are:

$$P(A, B) = \frac{p(A \cap B)}{p(A)p(B)}, \quad \text{for } A, B \in 2^X \backslash \varnothing \tag{29}$$

with p being an *a priori* probability measure assumed to exist on X representing some background knowledge on the hypotheses of X, such as class priors in a classification problem. The definition of (29) results in an extension of the "trivial plausibility measure" ρ defined such that $\rho(A) = 1$ if $A \neq \varnothing$ and $\rho(\varnothing) = 0$ which defines "Dempster's agreement" [26]. The resulting inner product is richer than $\otimes^s_I$ as it satisfies the strong structural property (sp1). For instance, $P(A, B) = 0$ if and only if $A \cap B = \varnothing$, otherwise it is between 0 and 1. The corresponding distance, denoted as $d^{(2)}_P$, is thus more appropriate than $d^{(2)}_I$ to quantify the distance between belief functions. However, $d^{(2)}_P$ is a pseudometric (the condition of separability (d3)" is not respected) since $\mathbf{P}$ is only positive semi-definite. This means that one can obtain $m_1 \neq m_2$ such

[9] We assume that the authors of [50] referred to Eq. (27) instead of their expression (34) which is always equal to 1, as noticed in [27].

that $d_P^{(2)}(m_1, m_2) = 0$. The definition of p is assumed to be known before hand and depends on the application at hand (for instance, prior distribution over a set of classes). However, without any knowledge about p a uniform distribution can be reasonably considered and we obtain $P(A, B) = \frac{N|A\cap B|}{|A||B|}$.

In [35], with the aim of defining a "full" distance accounting for the interaction between the focal elements (satisfies the strong structural property), we introduced an inner product based on the Jaccard index, itself a classical similarity measure between sets:

$$\otimes_J^s(m_1, m_2) = \mathbf{m}_1'\mathbf{Jac}\ \mathbf{m}_2 \tag{30}$$

where **Jac** is the matrix whose elements are Jaccard indexes:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}, \quad \text{for } A, B \in 2^X \setminus \varnothing \tag{31}$$

The matrix **Jac** is positive definite and the corresponding distance $d_J^{(2)}$ is a full metric.[10] Besides satisfying all the metric axioms,(d1) to (d4), $d_J^{(2)}$ also satisfies the strong structural property (sp1). This distance has been widely used in several applications such as for instance in the estimation of discounting rates [28,43,18,31].

In [24], Diaz et al. proposed to extend $d_J^{(2)}$ in two ways: by replacing $J(A, B)$ by any similarity measure between sets $S(A, B)$ and by using one of 17 possible measures for the definition of the weighting matrix. For instance, a Dice index can be used:

$$D(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \quad \text{for } A, B \in 2^X \setminus \varnothing \tag{32}$$

The properties of the resulting distance $d_S^{(2)}$ depend on the properties of the matrix **S** and the main remaining question is whether **S** is positive definite or semi-definite. Without a formal analysis of the properties of D, we cannot know if $d_D^{(2)}$ is or not a full metric. However, we will see in Section 5.2 how these properties can be suspected from the membership of $d_D^{(2)}$ to a particular family of measurements.

The other extension proposed in [24] (the main purpose of their work) relies on the modification of **S** by a function F so that the resulting similarity measure "rewards" small cardinalities while penalizing high cardinalities of focal sets. The F function is piecewise: One piece for "reward" and one piece for "penalty", and the connection is done through a ρ parameter. The modification of inner products is then defined by:

$$\otimes_{F(S)}^s(m_1, m_2) = \mathbf{m}_1'\ F(\mathbf{S}, R)\ \mathbf{m}_2 \tag{33}$$

where $R = \frac{|A\cup B|}{|X|}$, **S** is the matrix whose elements quantify the similarity between focal elements. As an example, let us consider $A = \{x_1, x_2, x_3, x_4, x_5\}$, $B = \{x_1, x_2, x_3\}$, $C = \{x_1, x_2\}$, $D = \{x_1\}$. Diaz et al. [24] argue that it is easier for A and B to be similar (since they have high cardinalities) than it is for C and D, and thus the similarity should be corrected accordingly. Hence, while we have $J(A, B) = J(C, D)$, we have $F(\mathbf{Jac}, R)(A, B) < F(\mathbf{Jac}, R)(C, D)$. All the distances based on (33) satisfy the strong structural property (sp1) as well as all the metric axioms excepted the separability axiom (d3)" depending on the definiteness of **S**.

With the aim of computing the orthogonal projection of a belief function onto the probability simplex, Cuzzolin proposed in [13] the standard Euclidean distance between belief functions. Although not defined explicitly, the underlying inner product is:

$$\otimes_{Inc}^s(m_1, m_2) = \mathbf{m}_1'\ \mathbf{Inc}\ \mathbf{Inc}'\ \mathbf{m}_2 \tag{34}$$

where **Inc** has been introduced in Section 2.2, and the elements of the matrix **Inc Inc**′ are:

$$\text{Inc}_2(A, B) = |\{A \subseteq C\} \cap \{B \subseteq C\}|, \quad \forall C \subseteq X \tag{35}$$

where $\{\cdot\}$ denotes a set of subsets of X. In other words, Inc_2 is the number of subsets C of X which contain both A and B. The weighting matrix **Inc Inc**′ is also a way to quantify the interaction between focal elements rather based on their inclusion than on their similarity. Thus, the resulting distance $d_{Inc}^{(2)}$ satisfies the strong structural property (sp1). Moreover **Inc Inc**′ is positive definite and the resulting distance, $d_{Inc}^{(2)}$, is thus a full metric. As already noticed, replacing the belief function by the plausibility, *i.e.* replacing **Inc** by **Int** leads to the same measure.

Another way to compare belief functions is through their betting ability: Two belief functions are close if their betting functions are close, *i.e.* if their pignistic transformations are close. Any distance between probability distributions can then be used. Zouhal and Denœux in [61] proposed to measure the distance between a belief function and an indicator vector using the L_2 measure based upon the inner product:

[10] Initially conjectured in [35], a formal proof of this property has been recently proposed in [9].

$$\otimes^{s}_{Betx}(m_1, m_2) = \mathbf{m}_1' \ \mathbf{Bet}_x' \ \mathbf{Bet}_x \ \mathbf{m}_2 \quad (36)$$

where $\mathbf{Bet}_x$ is the $N \times 2^N - 1$ matrix introduced in Section 2.2 and the elements of $\mathbf{Bet}_x' \ \mathbf{Bet}_x$ are:

$$\text{Bet}_{x2}(A, B) = \frac{|A \cap B|}{|A|.|B|}, \quad \text{for } A, B \in 2^X \backslash \varnothing \quad (37)$$

As previously noticed, $\mathbf{Bet}_x' \ \mathbf{Bet}_x$ is not of full rank, *i.e.* positive but only semi-definite, and the corresponding distance $d^{(2)}_{Betx}$ is a pseudometric. Nevertheless, $\mathbf{Bet}_x' \ \mathbf{Bet}_x$ quantifies the interaction between the focal elements and $d^{(2)}_{Betx}$ satisfies (sp1). Also, we have that $\text{Bet}_{x2}(A, B) = \frac{1}{N} P(A, B)$ whenever p is uniform over X, and $d^{(2)}_P = \sqrt{|X|} d^{(2)}_{Betx}$ thus $d^{(2)}_{Betx}$ has the same properties than $d^{(2)}_P$. Note also that the inner product (36) could have been defined through the square matrix **Bet**, as introduced at the end of Section 2.2 leading to a different distance $d^{(2)}_{Bet}$ but which remains proportional to $d^{(2)}_{Betx}$ as $d^{(2)}_{Bet} = \sqrt{2} d^{(2)}_{Betx}$. The distance $d^{(2)}_{Betx}$ has been used in [45] for learning discounting rates.

In [19], Denœux defines a series of Euclidean distances $d^{(2)}_v$ in the very special case of BPAs with $N + 1$ focal elements (N singletons plus X). Three distances defined in this work turn out to be $d^{(2)}_{Intx}$, $d^{(2)}_{Incx}$ and $d^{(2)}_{Betx}$.

3.2.3. Chebyshev distances – L_∞

Chebyshev distance is induced by the supremum (or infinity or uniform) norm and is equal to the limit of (20) when p grows toward $+\infty$. L_∞ relies on a max operator and with $p = \infty$, (20) becomes:

$$d^{(\infty)}_W(m_1, m_2) = \max_{A \subseteq X} \left\{ |(\mathbf{U} \ \mathbf{m}_1)' \mathbf{e}_A - (\mathbf{U} \ \mathbf{m}_2)' \ \mathbf{e}_A| \right\} \quad (38)$$

Aiming to assess the quality of Bayesian approximation algorithms of belief functions, Tessem [57] proposed three "error measures" which turn to belong to the L_∞ family of Chebyshev distances:

$$d^{(\infty)}_{Bet}(m_1, m_2) = \max_{A \subseteq X} \left\{ |(\mathbf{Bet} \ \mathbf{m}_1)' \ \mathbf{e}_A - (\mathbf{Bet} \ \mathbf{m}_2)' \ \mathbf{e}_A| \right\} \quad (39)$$

the equivalent measures between belief values:

$$d^{(\infty)}_{Inc}(m_1, m_2) = \max_{A \subseteq X} \left\{ |(\mathbf{Inc}' \ \mathbf{m}_1)' \ \mathbf{e}_A - (\mathbf{Inc}' \ \mathbf{m}_2)' \ \mathbf{e}_A| \right\} \quad (40)$$

and between plausibilities of singletons:

$$d^{(\infty)}_{Intx}(m_1, m_2) = \max_{x \in X} \left\{ |(\mathbf{Int} \ \mathbf{m}_1)' \ \mathbf{e}_x - (\mathbf{Int} \ \mathbf{m}_2)' \ \mathbf{e}_x| \right\} \quad (41)$$

where $\mathbf{e}_x$ is the singleton basis vector corresponding to x.

Note that L_∞ distances can be applied to continuous spaces in an easier way than L_1 and L_2 distances, hence a possible interest for continuous belief functions.

3.3. Inner product family

The general formulation for an inner product in $\mathcal{E}_X$ is:

$$\otimes_W(m_1, m_2) = \mathbf{m}_1' \mathbf{W} \mathbf{m}_2 \quad (42)$$

All the inner products introduced in Section 3.2 are summarized in Table 2. While they are all symmetric and linear regarding the first component (axioms (ip1) and (ip2)), some of them are degenerate as their matrix **W** is not positive-definite.

The weighting matrices **W** can be qualified as either a similarity or a dissimilarity matrix over 2^X, hence the corresponding superscript either s or d on $\otimes$.

3.3.1. Inner product

Dempster's conflict quantifies the conflict between two belief functions from two independent sources. Although it has not been defined to quantify a dissimilarity between two belief functions it has however been used for such a purpose as for instance in [52] or [23] in clustering algorithms and to some extent in [50] in association algorithms. It can be put under the form of an inner product as:

$$\otimes^{d}_{Int}(m_1, m_2) = \mathbf{m}_1' (1 - \mathbf{Int}) \ \mathbf{m}_2 \quad (43)$$

where **Int** is the matrix of intersections between two subsets of X introduced in Eq. (5). Note that $(1 - \mathbf{Int})$ is neither positive nor negative nor definite and should thus not be called an inner product but we put it in this family of measures as it satisfies the general formulation. We can also write:

Table 2
Inner products $\otimes_W(m_1, m_2) = \mathbf{m_1}'\mathbf{W}\mathbf{m_2}$. The superscript s of $\otimes$ means that $\mathbf{W}$ is a similarity measure between sets, whereas a superscript d means that $\mathbf{W}$ is a dissimilarity measure between sets.

Notation	W	Def.	Ref.
$\otimes_I^s$	$\mathbf{I}$	$I(A, B) = 1$ iff $A = B$	[49]
$\otimes_{Inc}^s$	$\mathbf{IncInc}'$	$\text{Inc}(A, B) = 1$ iff $A \subseteq B$	[14]
$\otimes_{Int}^s$	$\mathbf{Int}$	$\text{Int}(A, B) = 1$ iff $A \cap B \neq \emptyset$	[17]
$\otimes_{Int2}^s$	$\mathbf{Int}'\mathbf{Int}$	$\text{Int}(A, B) = 1$ iff $A \cap B \neq \emptyset$	[57]
$\otimes_{Intx}^s$	$\mathbf{Int}_x'\mathbf{Int}_x$	$\text{Int}_x(x, B) = 1$ iff $x \in B$	[57]
$\otimes_{Bet}^s$	$\mathbf{Bet}'\mathbf{Bet}$	$\text{Bet}(A, B) = \frac{\|A\cap B\|}{\|B\|}$	-
$\otimes_{Betx}^s$	$\mathbf{Bet}_x'\mathbf{Bet}_x$	$\text{Bet}_x(x, A) = \frac{\|x\cap A\|}{\|A\|}$	[61]
$\otimes_P^s$	$\mathbf{P}$	$P(A, B) = \frac{p(A\cap B)}{p(A)p(B)}$	[26]
$\otimes_J^s$	$\mathbf{Jac}$	$J(A, B) = \frac{\|A\cap B\|}{\|A\cup B\|}$	[35]
$\otimes_S^s$	$\mathbf{S}$	$S(A, B)$ any similarity measure	[24]
$\otimes_{F(S)}^s$	$F(\mathbf{S}, R)$	F reward-penalty function	[24]

$$\begin{aligned}\otimes_{Int}^d(m_1, m_2) &= 1 - \mathbf{m}_1'\mathbf{Int}\ \mathbf{m}_2 \\ &= 1 - \otimes_{Int}^s(m_1, m_2)\end{aligned} \tag{44}$$

where $\otimes_{Int}^s(m_1, m_2)$ is *Dempster's agreement* as called in [26]. We can easily check that **Int** defines a similarity measure over 2^X (Axioms (s1), (s2) and (s3) are satisfied) whereas 1 — **Int** defines a dissimilarity measure over 2^X (Axioms (d1), d(2) and (d3)' satisfied in 2^X). Unfortunately, this does not imply that $\otimes_{1-Int}^s$ is a dissimilarity measure in $\mathcal{E}_X$ as it is indeed not the case. We note that $\otimes_{Int}^d = \otimes_{1-Int}^d = 1 - \otimes_{Int}^s$.

The inadequacy of $\otimes_{Int}^d$ to characterize the dissimilarity between belief functions has been noticed in several works, as for instance in [41] or in [43] where the authors highlight the fact that the internal conflict $\otimes_{Int}^d(m, m)$ is not 0. We can easily check that $\otimes_{Int}^d$ indeed satisfies (d1) and (d2) but fails to satisfy the reflexivity and separability axioms (d3)' and (d3)'' and thus $\otimes_{Int}^d$ is a nonmetric measure. $\otimes_{Int}^d$ however satisfies the strong structural property (sp1). Fig. 6 in Section 5.4 illustrates $\otimes_{Int}^d$'s behavior compared to a known dissimilarity measure.

Based on $\otimes_{Int}^d$, Ristic and Smets [50] defined what they called an "additive global dissimilarity measure" as:

$$d_{RS}(m_1, m_2) = -\log\left(1 - \otimes_{Int}^d(m_1, m_2)\right) \tag{45}$$

Although d_{RS} has the properties of $\otimes_{Int}^d$, it suffers from the same drawback than $\otimes_{Int}^d$ and does not satisfy the axiom (d3)' of reflexivity, *i.e.* $d_{RS}(m, m) \neq 0$. Moreover, d_{RS} is not defined whenever $\otimes_{Int}^d(m_1, m_2) = 1$.

3.3.2. Cosine

The cosine measure however, defines a measure of similarity between belief functions. A general formulation is given by:

$$\cos_W(m_1, m_2) = \frac{\mathbf{m}_1'\mathbf{W}\ \mathbf{m}_2}{||\mathbf{m}_1||_W \cdot ||\mathbf{m}_2||_W} \tag{46}$$

In [59], Wen et al. use $\otimes_I^s = \mathbf{m}_1'\mathbf{I}\ \mathbf{m}_2$ to define a cosine measure resulting in a valid measure of similarity between two belief functions which, as $\otimes_I^s$, does not satisfy any structural property. A distance measure in $\mathcal{E}_X$ with semipseudometric properties can be obtained by:

$$\cos_W^d(m_1, m_2) = 1 - \cos_W(m_1, m_2) \tag{47}$$

Indeed, $\cos_W^d$ satisfies Axioms (d1), (d2) and (d3)' as soon as $\mathbf{W}$ is positive semidefinite. More particularly, (d2) and (d3)' are always satisfied whatever the properties of $\mathbf{W}$, but (d1) is satisfied only if $\mathbf{W}$ is positive.

3.4. Fidelity family

Fidelity is a popular measure of distance in quantum theory, and is based on the square root of probability distributions. The general formulation of the fidelity coefficient, also known as the Bhattacharyya coefficient, for belief functions is given by [6]:

$$\otimes_W^{\frac{1}{2}}(m_1, m_2) = \sqrt{\mathbf{m}_1}'\mathbf{W}\ \sqrt{\mathbf{m}_2} \tag{48}$$

where $\sqrt{\mathbf{m}} = (\mathbf{m}_1)^{\frac{1}{2}}$ is the vector obtained by taking the square roots of each component of $\mathbf{m}$. All measures of this family are functions of some $\otimes_W^{\frac{1}{2}}$. Fidelity family is also called "squared-chord" family and is a popular distance with paleontologists and in palynology studies. In probability theory, the Bhattacharyya (or fidelity) coefficient measures the amount of overlap between two statistical populations.

It has been originally extended to belief functions by Ristic and Smets in [50] who proposed a Hellinger distance [11] [33] based on $\otimes_W^{\frac{1}{2}}$:

$$d_I^{(H)}(m_1, m_2) = \left(1 - \otimes_I^{\frac{1}{2}}(m_1, m_2)\right)^{\frac{1}{2}} \tag{49}$$

A modified version of (49) has been proposed by Florea and Bossé in [27], obtained by replacing the square root by any a-root, with $a \in \mathbb{R}^{+*}$. Based on $\mathbf{I}$, this distance does not satisfy either (sp1) nor (sp2) but it is nevertheless a full metric (Axioms (d1) to (d4) are satisfied).

3.5. Information-based distances

Besides the extension of the probabilistic form of the metric distance family, another quantification of the idea of distance between two belief functions can be materialized by estimating the difference in their information content. As introduced at the beginning of this section, the first attempt is due to Perry and Stephanou [46] who proposed an extension of Kullback–Liebler divergence. But other works are worth mentioning.

In [20], in order to measure the quality of belief function approximations, Denœux proposed to quantify the distance between m_1 and m_2 by the difference between the information contents of m_1 and m_2. The measure of uncertainty used is the generalized cardinality of a belief function defined in [25] as $GC(m) = \sum_{A \subseteq X} m(A)|A|$ or using the matrix notation, $GC(m) = \mathbf{m}'\mathbf{c}_A$ where $\mathbf{c}_A$ is the column vector of cardinalities [12] of A (denoted as **CardA** in [53]). We have thus:

$$d_{GC}(m_1, m_2) = |(\mathbf{m}_1 - \mathbf{m}_2)'\mathbf{c}_A| \tag{50}$$

The general formulation was also mentioned in [20]:

$$d_U(m_1, m_2) = |U(m_1) - U(m_2)| \tag{51}$$

where U is any uncertainty measure defined for belief functions (see for instance [39] for a survey). Note that for the practical use in [20] it was always true that $GC(m_1) \geq GC(m_2)$ and thus that $d_{GC}(m_1, m_2) \geq 0$. Although the original measure did not explicitly include the absolute value we think it was implicit. And accordingly we have decided to add it in the present definition insuring that the resulting distance satisfies the minimal property of nonnegativity (axiom (d1)). Moreover, with the formulation (50) axioms (d2) and (d3)' are satisfied while axioms (d3)" and (d4) are not satisfied, making d_{GC} a semipseudometric.

Also, in [19], Denœux proposed that a cross-entropy measure $d_{CE}(m_1, m_2) = -\mathbf{m}_1' \log_2(\mathbf{m}_2) + (1 - \mathbf{m}_1)' \log_2(1 - \mathbf{m}_2)$ could be used as an alternative to the Euclidean distance $d_v^{(2)}$ for optimizing neural network weights.

3.6. Two-dimensional distances

In [41], Liu defined a two-dimensional measure to formally quantify the conflict between belief functions, as she noticed that neither Dempster's conflict alone, nor a distance alone is satisfactory. She then proposed the following index:

$$d_L^{2D} = \left(\otimes_{Int}^d(m_1, m_2); d_{Bet}^{(\infty)}(m_1, m_2)\right) \tag{52}$$

She argues that "only when both measures are high, it is safe to say the evidence is in conflict" [41]. Indeed, $\otimes_{Int}^d$ acts as the angle measure while $d_{Bet}^{(\infty)}$ is the metric measure, the two measures being based on two different inner products (see Section 4.5). This principle can be extended to any other two inner products and the general formulation extending (52) is then:

$$d_{(W,V)}^{2D}(m_1, m_2) = \left(\otimes_W^d(m_1, m_2); d_V(m_1, m_2)\right) \tag{53}$$

It is not required for $\otimes^d$ and d to be defined upon the same inner product, hence one may be based on a weighting matrix $\mathbf{W}$ while the other on a different matrix $\mathbf{V}$, possibly increasing the complementarity of the two measures. As we will see in Section 5.2, the correlation coefficient between measures may also be a criterion for building two-dimensional measures.

4. Outcomes

We provided four general formulations of distances between belief functions in Eqs. (20), (42), (47) and (48), which encompass most of the distances defined so far in the technical literature. Moreover, we provided a general formulation (53)

[11] Up to a factor 2.

[12] Vector $\mathbf{c}_A$ is closely linked to vector $\mathbf{g}_A$ introduced in Section 3.1 and we have $\mathbf{g}_A = \frac{\mathbf{c}_A - 1}{|X|-1}$.

Table 3
Distances between belief functions in their respective family (L_p, inner product or Fidelity), according to several definitions of the weighting matrix **W**. The distances defined so far are in gray cells while new ones are in white cells.

	L_p			Inner product		Fidelity
$\mathbf{W} = \mathbf{U}'\mathbf{U}$	$p = 1$	$p = 2$	$p = \infty$	IP	cos	(Hellinger)
I	$d_I^{(1)}$	$d_I^{(2)}$	$d_I^{(\infty)}$	$\otimes_I^s$	$\cos_I$	$d_I^{(H)}$
IncInc′	$d_{Inc}^{(1)}$	$d_{Inc}^{(2)}$	$d_{Inc}^{(\infty)}$	$\otimes_{Inc}^s$	$\cos_{Inc}$	$d_{Inc}^{(H)}$
Int	$d_{Int}^{(1)}$	$d_{Int}^{(2)}$	$d_{Int}^{(\infty)}$	$\otimes_{Int}^s$	$\cos_{Int}$	$d_{Int}^{(H)}$
Int′Int	$d_{Int2}^{(1)}$	$d_{Int2}^{(2)}$	$d_{Int2}^{(\infty)}$	$\otimes_{Int2}^s$	$\cos_{Int2}$	$d_{Int}^{(H)}$
$\mathbf{Int}_x'\mathbf{Int}_x$	$d_{Intx}^{(1)}$	$d_{Intx}^{(2)}$	$d_{Intx}^{(\infty)}$	$\otimes_{Intx}^s$	$\cos_{Intx}$	$d_{Intx}^{(H)}$
Bet′Bet	$d_{Bet}^{(1)}$	$d_{Bet}^{(2)}$	$d_{Bet}^{(\infty)}$	$\otimes_{Bet}^s$	$\cos_{Bet}$	$d_{Bet}^{(H)}$
$\mathbf{Bet}_x'\mathbf{Bet}_x$	$d_{Betx}^{(1)}$	$d_{Betx}^{(2)}$	$d_{Betx}^{(\infty)}$	$\otimes_{Betx}^s$	$\cos_{Betx}$	$d_{Betx}^{(H)}$
P	$d_P^{(1)}$	$d_P^{(2)}$	$d_P^{(\infty)}$	$\otimes_P^s$	$\cos_P$	$d_P^{(H)}$
Jac	$d_J^{(1)}$	$d_J^{(2)}$	$d_J^{(\infty)}$	$\otimes_J^s$	$\cos_J$	$d_J^{(H)}$
S	$d_S^{(1)}$	$d_S^{(2)}$	$d_S^{(\infty)}$	$\otimes_S^s$	$\cos_S$	$d_S^{(H)}$
$F(\mathbf{S}, R)$	$d_{F(S,R)}^{(1)}$	$d_{F(S,R)}^{(2)}$	$d_{F(S,R)}^{(\infty)}$	$\otimes_{F(S,R)}^s$	$\cos_{F(S,R)}$	$d_{F(S,R)}^{(H)}$

Table 4
Axiomatic properties of the distances defined so far (see Table 1 and Section 2.5 for the axiom definitions).

Distance	Metric					Structural		
	(d1)	(d2)	(d3)’	(d3)”	(d4)	(sp1)	(sp2)	(sp3)
$d_I^{(H)}$	×	×	×	×	×			
$d_I^{(2)}$	×	×	×	×	×			
$d_{Inc}^{(2)}$	×	×	×	×	×	×		
$d_J^{(2)}$	×	×	×	×	×	×		
$d_{F(J)}^{(2)}$	×	×	×	×	×	×		
$d_D^{(2)}$	×	×	×	×	×	×		
$d_{Betx}^{(2)}$	×	×	×		×	×		
$d_{Bet}^{(\infty)}$	×	×	×		×	×		
$d_P^{(2)}$	×	×	×		×	×		
d_{GC}	×	×	×				×	
$\cos_I^d$	×	×	×					
$\otimes_{Int}^d$	×	×				×		
d_{RS}	×	×				×		
d_{BP}		×				×	×	
$d_{BP(1)}$	×	×				×		
$d_{BP(2)}$		×					×	
d_{PS}		×				×		×
$d_{PS(1)}$	×	×						×
$d_{PS(2)}$		×				×		

for a two-dimensional measure. In this section we synthesize further our survey and sketch some ideas to be developed in future research.

4.1. Summary and new measures

Table 3 summarizes the distances defined so far (gray cells) and provides the natural generalizations (white cells) hence new distances.

Twenty-three distances have been defined so far and the generalization has led to more than 40 new ones. The Minkowski L_2 family is the most numerous while L_1 and L_∞ have been seldom used. Extending the L_2 distances to the study of L_1 and L_∞ distances requires in some cases a Cholesky decomposition of the weighting matrices **W**. A single cosine measure has been defined so far but a multitude of measures of this kind remains to be explored. This comment also applies to the Hellinger distance and to other distances of the Fidelity family which are based on Bhattacharyya’s coefficient involving the squared-root of the BPAs.

4.2. Metric and structural properties

Table 4 summarizes the algebraic properties of the distances together with their structural properties.

Distances with the highest number of metric properties appear at the top of the table while weakest metric distances appear at the bottom. All distances are symmetric. Most of the distances are nonnegative. However, the two composite measures d_{PS} and d_{BP} may have negative values due in both cases to their second member $d_{PS(2)}$ and $d_{BP(2)}$ respectively. The second discrimination criterion between the distances is the definiteness property (d3) which splits into the reflexivity property (d3)’ and the separability property (d3)”. The reflexivity means that $d(m, m) = 0$ and if not satisfied makes the distances qualifying as nonmetric distances. In our case of study, distances not satisfying this property are all based on

Table 5
Classes of distances according to their metric properties crossed with the structural properties. The (★) symbol on d_{GC} indicates that it satisfies the weak structural property (sp2) while the others all satisfy the strong structural property (sp1).

	Metric	Pseudometric	Semipseudometric	Nonmetric
Structural	$d_D^{(2)}$ $d_{F(J)}^{(2)}$ $d_J^{(2)}$ $d_{Inc}^{(2)}$	$d_{Betx}^{(2)}$ $d_{Bet}^{(\infty)}$ $d_P^{(2)}$	d_{GC} (★)	$\otimes_{Int}^d$ d_{RS} d_{BP} d_{PS}
Non-structural	$d_I^{(2)}$ $d_I^{(H)}$		$\cos_I^d$	

Table 6
Ranges of the distances.

	Min.	Max.
$d_I^{(H)}$	0	1
$d_I^{(2)}$	0	$\sqrt{2}$
$d_{Inc}^{(2)}$	0	$\sqrt{2^{\lvert X\rvert-1}}$
$d_J^{(2)}$	0	$\sqrt{2}$
$d_{F(J)}^{(2)}$	0	$\sqrt{2}$
$d_D^{(2)}$	0	$\sqrt{2}$
$d_{Bet}^{(2)}$	0	$\sqrt{2}$
$d_{Bet}^{(\infty)}$	0	1
$d_P^{(2)}$	0	$\sqrt{2\lvert X\rvert}$
d_{GC}	0	$\lvert X\rvert - 1$
$\cos_I^d$	0	1
$\otimes_{Int}^d$	0	1
d_{RS}	0	$+\infty$
d_{BP}	-1	$+\infty$
d_{PS}	-1	$2^{\lvert X\rvert}$

Dempster's conflict or Dempster's rule. The reflexivity property is often desirable as intuitively, two identical belief functions should have a distance of 0. The separability property guaranties that a null value for the distance implies that it has been computed between two identical belief functions. Only 6 distances satisfy (d3)". To be able to conclude on the practical utility of such a property a case by case study would be required since the possible situation of two distinct belief functions yielding a null distance may not have a high impact in some practical situations. The last property is the triangle inequality (d4) which guaranties that the direct path between two belief functions is always lower than when a third belief function is involved. Among the metric distances, only d_{GC} and $\cos_I^d$ do not satisfy (d4) which make them semipseudometrics.

Three distances do not satisfy any structural property, namely $d_I^{(2)}$, $d_I^{(H)}$ and $\cos_I^d$, all based obviously on the identity matrix. d_{GC} satisfies a weak form of structural property (sp2) and d_{PS} satisfies two structural properties (sp1) and (sp3), the latter property appearing to be even weaker than (sp2). The strong structural property (sp1) is satisfied by all the other distances. Structural properties are desirable when quantifying the distance between belief functions as they allow to distinguish the standard distances between probability distributions from the specific distances between belief functions. Indeed, relaxing the additivity axiom does not shelter us from the particular feature of the space of belief functions whose basis components remain linked by some underlying similarity measures or at least some interactions.

Eight (8) classes [13] have been identified and shown in Table 5 regarding the properties of Table 4, *i.e.* full metrics, pseudometrics, semipseudometrics and nonmetric measures cross-tabulated with the structural properties.

The most two populated classes are the structural metrics (all belonging to the L_2 family) and the structural nonmetrics. We have only two distinct cases of pseudometrics (since $d_{Betx}^{(2)} \propto d_P^{(2)}$), based on the same weighting matrix and belonging to the Minkowski family. The single structural semipseudometric d_{GC} satisfies only the weak structural property. Although most of classes have at least one candidate, some are empty and in particular we have no example of nonstructural pseudometric nor nonmetric. Four distances namely $d_{Inc}^{(2)}$, $d_J^{(2)}$, $d_{F(J)}^{(2)}$ and $d_D^{(2)}$ qualify as structural metrics and satisfy thus the highest number of properties considered in this work. The experimental study of distances' behaviors of Section 5.2 will be structured according to these classes.

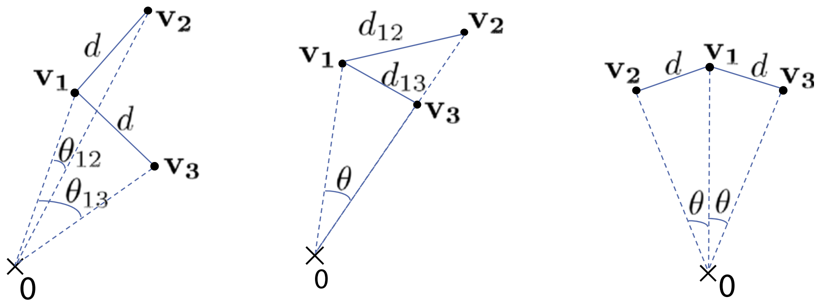
4.3. Normalization

All the distance measurements introduced in Section 3 have been presented without any normalization factor. In Table 6, the range of each distance is computed, these values could possibly be used for normalizing the distances.

In accordance to the metric properties, all distances have a minimum value of 0, except d_{PS} and d_{BP} which can have negative values. [14] Two distances are not bounded, d_{RS} and d_{BP} due to the use of the log function. Most of the maximum values are constant while 4 depend on $\lvert X\rvert$.

---

[13] We reserve the term "family" for categorizing the distances according to their definitions (L_2, cosine, Fidelity, etc.) while the term "class" is used for categorizing the distances according to their axiomatic properties (full metric, pseudometric, etc.).
[14] Note that the exact minimum value has not been computed and may been higher than -1 while being obviously negative.



(a) Equal distances and different angles.

(b) Equal angles and different distances.

(c) Equal angles and equal distances.

Fig. 2. Two dimensions for the measures of dissimilarity between belief functions.

4.4. Bayesian belief functions

The property which guaranties that a distance between belief functions reduces to a known distance between probability distributions whenever the belief functions are Bayesian is called "probability consistency". Among the distances identified, all the Minkowski distances reduce to the traditional Minkowski distances between probability distributions $d(p_1, p_2) = \sqrt[p]{|\mathbf{p}_1 - \mathbf{p}_2|^p}, p = 1, 2, +\infty$. The cosine measure remains the cosine measure restricted to the probability space. Dempster's conflict is $1 - \mathbf{p}_1'\mathbf{p}_2$ and corresponds to 1 minus the inner product between two probability distributions, $d_I(H)$ reduces to a Hellinger distance. d_{PS}, d_{BP}, d_{RS} d_{GC} reduce to no known standard measures.

4.5. Metric and angle as complementary measures

At least two kinds of measures quantifying a notion of "distance" between three vectors $\mathbf{v}_1, \mathbf{v}_2$ and $\mathbf{v}_3$ exist: (1) the metric which quantifies *how close* two vectors are from each others ($d(\mathbf{v}_1, \mathbf{v}_2) < d(\mathbf{v}_1, \mathbf{v}_3)$ means that $\mathbf{v}_2$ is closer to $\mathbf{v}_1$ than $\mathbf{v}_3$ is (see Fig. 2(b))); (2) the angle which quantifies *how orthogonal* two vectors are ($\cos(\mathbf{v}_1, \mathbf{v}_2) < \cos(\mathbf{v}_1, \mathbf{v}_3)$ means that $\mathbf{v}_2$ is more parallel (or less orthogonal) to $\mathbf{v}_1$ than $\mathbf{v}_3$ is (see Fig. 2(a))). Two vectors can thus be very far while collinear (angle is null), whereas they can be very close while being orthogonal. Fig. 2 illustrates three cases of interaction between three vectors: equal distances and different angles, equal angles and different distances and equal angles and equal distances. There exist several pairs of vectors which have the same angle θ (or cosine value) and distinct distances, as in Fig. 2(a), the pair (cos; d) does not allow to completely discriminate the interaction between $\mathbf{v}_i$s as it defines a kind of "cone" around a symmetry axis (see Fig. 2(c)).

Metric and cosine measures can thus be seen as complementary measures of distance between belief functions. In the survey of Section 3, the pair $(\cos_I^d, d_I^{(2)})$, where $\cos_I^d = 1 - \cos_I$, characterizes globally how much two belief functions are far and orthogonal. This pair can be generalized to any other weighting matrix $\mathbf{W}$, and also to other pairs $(\cos_W^d, d_V)$. Thus, we have a general pair:

$$d^{2D}_{(W,V)}(m_1, m_2) = \left(\cos^d_W(m_1, m_2); d_V(m_1, m_2)\right) \tag{54}$$

where the inner product $\otimes^d$ of (53) has been replaced by the cosine measure $\cos^d$ in (54). Compared to the 2D measure initiated by Liu [41], we change the inner product by the cosine measure because the latter has more metric properties than the former, in particular it satisfies the reflexivity axiom (d3)'. This point will be developed in the following section in the specific case of Dempster's conflict.

4.6. Alternative to Dempster's conflict

Whereas the metric distance is a natural measure of dissimilarity, the cosine measure is itself a natural measure of similarity. Indeed, if θ denotes the angle between two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, then $\cos(\theta) = -1$ means that $\mathbf{v}_1$ and $\mathbf{v}_2$ are opposite ($\mathbf{v}_1 = -\alpha\mathbf{v}_2$, with $\alpha > 0$), $\cos(\theta) = 1$ means that $\mathbf{v}_1$ and $\mathbf{v}_2$ are the collinear ($\mathbf{v}_1 = \alpha\mathbf{v}_2$), $\cos(\theta) = 0$ means that they are orthogonal and in between values represent intermediate similarity values. In the specific case of belief functions, $\mathbf{v} = \mathbf{m}$ and since $m(A) \geq 0$ for all $A \subseteq X$, we always have that $0 \leq \cos(\theta) \leq 1$.

Dempster's conflict defines a notion of orthogonality in the sense that two belief functions are orthogonal (according to Dempster) if their conflict is 1. However, as noticed in Section 3.3, Dempster's conflict $\otimes^d_{Int}$ is even not a premetric as it fails to satisfy the reflexivity axiom (d3)', an essential property in many applications.

In order to maintain the notion of conflict that it defines while having the properties of a semipseudometric, one could think of the following cosine-based measure:

$$\cos^d_{Int}(m_1, m_2) = 1 - \frac{\mathbf{m}_1' \mathbf{Int} \mathbf{m}_2}{\|\mathbf{m}_1\|_{Int} \cdot \|\mathbf{m}_2\|_{Int}} \tag{55}$$

$$= 1 - \frac{\otimes^s_{Int}(\mathbf{m}_1, \mathbf{m}_2)}{\sqrt{\otimes^s_{Int}(\mathbf{m}_1, \mathbf{m}_1)}.\sqrt{\otimes^s_{Int}(\mathbf{m}_2, \mathbf{m}_2)}} \tag{56}$$

where $\|\cdot\|_{Int}$ denotes the norm relatively to the matrix **Int** and $\otimes^s_{Int}$ is Dempster's agreement. Indeed, normalizing the inner product by both the norm of $\mathbf{m}_1$ and $\mathbf{m}_2$ guaranties that $\cos^d_{Int}(m, m) = 0$, axiom (d3)' of reflexivity is satisfied. Unfortunately, because **Int** is nonpositive, $\cos^d_{Int}$ is not positive. So we gained one interesting property but we lost another one, the nonnegativity (d1).

The alternative matrix **Int Int** however is positive definite and could thus be used as a basis for an alternative measure to Dempster's conflict:

$$\cos^d_{Int2}(m_1, m_2) = 1 - \frac{\mathbf{m}_1' \mathbf{Int}\ \mathbf{Int} \mathbf{m}_2}{\|\mathbf{m}_1\|_{Int2} . \|\mathbf{m}_2\|_{Int2}} \tag{57}$$

$$= 1 - \frac{\mathbf{Pl}_1' \mathbf{Pl}_2}{\|\mathbf{Pl}_1\| . \|\mathbf{Pl}_2\|} \tag{58}$$

where we noticed that **Int** is symmetric and $\mathbf{Pl} = \mathbf{Int}\ \mathbf{m}$. Now the three basic axioms of a semipseudometric measure (d1), (d2) and (d3)' are satisfied. But remains the question of the meaning of **IntInt**. Indeed, the elements of **IntInt** are:

$$\mathrm{Int}_2(A, B) = |\{A \cap C\} \cap \{B \cap C\}|, \forall C \subseteq X \tag{59}$$

that is the number of subsets $C \subseteq X$ intersecting with both A and B. $\cos^s_{Int2}$ quantifies thus a "second-order" notion of agreement since Int_2 quantifies how much two given subsets intersect through the mediation of a third one. The "first-order" agreement (Dempster's) only quantifies if or if not two given subsets intersect directly with each other. The minimum agreement between two subsets A and B according to $\mathrm{Int}_2(A, B)$ is 2^{N-2} and is reached when $A \cap B = \emptyset$ and A and B are singletons. If $A \cap B = \emptyset$ and A and B are not singletons, then Int_2 is thus higher and depends on the cardinalities of A and B. The maximum of $\mathrm{Int}_2(A, B)$ is $\mathrm{Int}_2(X, X) = 2^N - 1$. Thus, two belief functions without focal elements in interaction, which would mean a null agreement according to $\otimes^s_{Int}$ may exhibit a nonnull agreement according to $\cos^s_{Int2}$.

The associated Euclidean distance in the inner product space with **Int Int** as weighting matrix is:

$$d^{(2)}_{Int2}(m_1, m_2) = \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)' \mathbf{Int}' \mathbf{Int} (\mathbf{m}_1 - \mathbf{m}_2)} \tag{60}$$

$$= \|\mathbf{Pl}_1 - \mathbf{Pl}_2\| \tag{61}$$

$$= \|\mathbf{Bel}_1 - \mathbf{Bel}_2\| \tag{62}$$

$$= d^{(2)}_{Inc} \tag{63}$$

Thus, the standard Euclidean distance between belief functions would be the associated metric to a measure of angle derived from a "second-order" Dempster's conflict. This is due to the duality of Pl and Bel in the closed world. Nonetheless, the two weighting matrices **Int Int** and **Inc Inc**′ are different and thus $\cos^d_{Int2} \neq \cos^d_{Inc}$ and they measure two distinct kinds of angles.

4.7. Encoding of belief functions

The four well known functions of basic probability assignment, belief, plausibility and commonality are four different encodings of the same information. Defining Euclidean distances between two functions of the same kind leads to different distances when referring to a representation of reference that is a BPA, hence different values for the same two objects. For instance, we saw that $d^{(2)}_I$ (the Euclidean distance between BPAs) is different from $d^{(2)}_{Inc}$ (the Euclidean distance between belief functions). Also we have $d^{(2)}_{Inc} = d^{(2)}_{Int2}$ due to the duality between Pl and Bel under the closed world assumption. But $d^{(2)}_{Inc'}$, the Euclidean distance between two commonality functions built upon the weighting matrix **Inc**′**Inc** is different from the three others and would be worth to be considered in future works.

Besides these well known encodings of belief functions other encodings can be defined with however less obvious interpretations. For instance, we may define:

$$\mathbf{f}_J = \mathbf{U}_J \mathbf{m} \tag{64}$$

where $\mathbf{U}_J$ is the upper triangle matrix resulting from the Cholesky decomposition of the Jaccard matrix **Jac**. Or expressed under the form of Eq. (10):

$$f_J(A) = \sum_{B \subseteq X} m(B) U_J(A, B) \tag{65}$$

where $U_J(A, B)$ is the element (A, B) of matrix $\mathbf{U}_J$. Equivalently, we could define f_D, f_P, etc, corresponding respectively to Dice matrix, BPAM matrix.

The interest of considering the basic encoding for belief functions (*i.e.* the BPA) lies in the fact that it highlights the strong structural property of the weighted Euclidean distance. However, we could define some combinations such as for instance Bel_J defined such that $\mathbf{Bel}_J = \mathbf{U}_J\mathbf{Bel} = \mathbf{U}_J\mathbf{Inc}'\mathbf{m}$. Then the Euclidean distance between Bel_J^1 and Bel_J^2 would be defined as:

$$\begin{aligned} d_I^{(2)}(\text{Bel}_J^1, \text{Bel}_J^2) &= d_{IncJ}^{(2)}(m_1, m_2) \\ &= \sqrt{(\mathbf{Bel}_J^1 - \mathbf{Bel}_J^2)'(\mathbf{Bel}_J^1 - \mathbf{Bel}_J^2)} \\ &= \sqrt{[\mathbf{U}_J\mathbf{Inc}'(\mathbf{m}_1 - \mathbf{m}_2)]'\mathbf{U}_J\mathbf{Inc}'(\mathbf{m}_1 - \mathbf{m}_2)} \\ &= \sqrt{(\mathbf{m}_1 - \mathbf{m}_2)'\mathbf{Inc}\ \mathbf{Jac}\ \mathbf{Inc}'(\mathbf{m}_1 - \mathbf{m}_2)} \end{aligned} \tag{66}$$

Such formulations would reinforce the impact of the similarity between the basis vectors of $\mathcal{E}_X$.

4.8. Generalizations

1. Fuzzy belief functions:
   The above measures can be further generalized to fuzzy belief functions by making the weights $W(A, B)$ be measures of similarity between fuzzy sets. As an example, in [47] Petit-Renaud and Denœux defined an error criterion for fuzzy belief structures based on $\otimes_{PR}^{d}(m_1, m_2) = \mathbf{m}_1'\mathbf{D}_{PR}\mathbf{m}_2$ where $\mathbf{D}_{PR}$ is the matrix composed of elements $d_{PR}(\widetilde{A}, \widetilde{B})$ which itself is a distance between fuzzy sets $\widetilde{A}$ and $\widetilde{B}$.
2. Open world assumption:
   We restricted the study of the distances to the closed world assumption (or normality assumption, *i.e.* $m(\emptyset) = 0$). Although the extension to the open world assumption would require a deeper analysis, a preliminary analysis showed that most of the distances presented keep all their properties for BPAs with a non-null mass to the empty set, while others degenerate from metrics to pseudometrics. The latter are L_p distances involving weighting matrices which become nondefinite when the dimension $\mathbf{e}_\emptyset$ is added ($d_J^{(p)}$, $d_{F(J)}^{(p)}$, $d_D^{(p)}$). It seems that $d_{Inc}^{(p)}$ keeps its definiteness property.

5. Experimental comparison

5.1. Toy example

We illustrate the behavior of the different classes of distances of Table 5 on the toy example introduced in [35] and reused in [41]. Let $X = \{x_1, \ldots, x_{10}\}$ be a frame of discernment and let Bel_t be a belief function defined over X such that

$$\text{Bel}_t = \{(X, 0.1), (\{x_2, x_3, x_4\}, 0.05), (\{x_7\}, 0.05), (A_t, 0.8)\} \tag{67}$$

A_t is a variable focal element ranging from $\{x_1\}$ to X, one element x_i being added at each step. Let Bel^* be a categorical belief function representing a targeted belief function (*e.g.* representing the ground truth in a given problem) and defined by

$$\text{Bel}^* = (\{x_1, x_2, x_3, x_4, x_5\}, 1) \tag{68}$$

The three graphics in Fig. 3 show the behaviors of the distances identified in Section 3 assembled into the several families identified in Table 5, *i.e.* (full) metrics in the first graphic, pseudometrics and semipseudometrics in the second graphic, and nonmetrics in the third graphic. Nonstructural distances are represented with dashed lines. In each of the three graphics, distances of a variable belief function Bel_t (one focal element A_t only is variable) to a categorical belief function Bel^* of the form $m^*(\{x_1, x_2, x_3, x_4, x_5\}) = 1$ are shown, with A_t varying from $\{x_1\}$ to $\{x_1, \ldots, x_{10}\}$. For the simulations, all distances have been normalized with the range values of Table 6.

For the metrics class of distances (first graphic), we observe similar behaviors for all the structural distances while denoting however a small difference for the $d_{Inc}^{(2)}$ distance. The two nonstructural distances behave similarly while differently from the structural ones, as they unsurprisingly remain constant when A_t varies and only decrease when A_t reaches A^* at time step 5. We also note a slight increase of $d_I^{(2)}$ at the last time step when A_t reaches X since the number of focal elements for Bel_t has changed from 4 to 3.

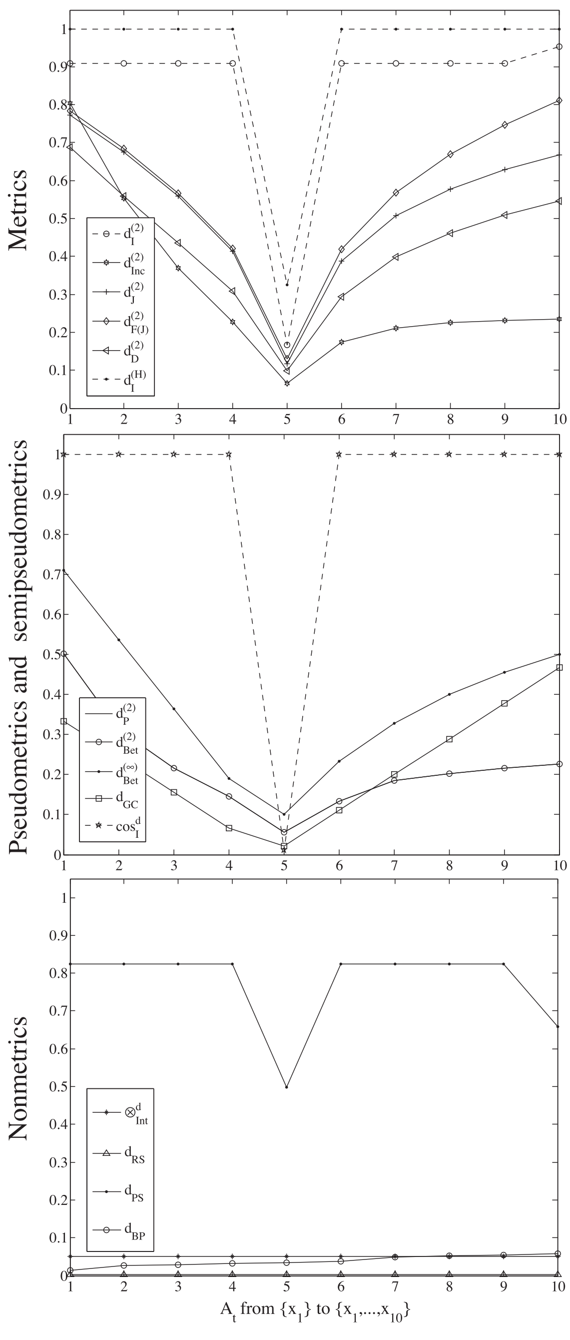


Fig. 3. Distances of a variable belief function Bel_t (*i.e.* A_t varies from $\{x_1\}$ to $\{x_1, \ldots, x_{10}\}$) to a categorical belief function Bel^*, *i.e.* $m^*(\{x_1, x_2, x_3, x_4, x_5\}) = 1$. Nonstructural distances are represented with dashed lines.

The second graphic shows the behaviors of the pseudometrics and the semipseudometrics together. Only 3 pseudometrics are visible since $d_{Betx}^{(2)} \propto d_P^{(2)}$ and because the distances have been normalized, $d_P^{(2)}$ and $d_{Betx}^{(2)}$ are confounded. The two pseudometrics of type d_{Bet} (and of course $d_P^{(2)}$) behave similarly. The two semipseudometrics have distinct behaviors: d_{GC} increases linearly with the number of elements in A_t, while the nonstructural cosine measure remains constant while A_t varies and reaches a minimum at time step 5 where A_t reaches A^*.

The nonmetric measures are shown in the third graphic. The two belief functions considered yield to a constant Dempster's conflict, $\otimes_{Int}^d(m_t, m^*) = m_t(\{x_7\})m^*(\{x_1, x_2, x_3, x_4, x_5\}) = 0.05$, and consequently d_{RS} and d_{BP} remain also very low since they are based on $\otimes_{Int}^d$. We observe however a slight increase of d_{BP} due to its second member $d_{BP(2)}$. The behavior of d_{PS} is similar to other nonstructural distances while we note a decrease at the last time step when $A_t = X$. Indeed, $d_{PS(1)}$, the first member of d_{PS}, remains constant and equal to 5 except when A_t reaches A^* where it is 3. The second member $d_{PS(2)}$ remains negative [15].

In the light of the example above, it appears that the structural properties (sp1) and (sp2) strongly influence the behavior of the different distances. Moreover, the structural dissimilarity (sp3) (only satisfied by d_{PS}) is not sensitive to this example of convergence toward a categorical belief function, and makes this "structural distance" close to nonstructural ones. The accounting of the focal elements interaction defined by (sp3) (*i.e.* the interaction of sets of focal elements) is thus too rough compared to (sp1) and (sp2). Also, we do not denote any significant behavior difference between metrics and pseudometrics while distances involving Dempster's conflict behave similarly.

5.2. Semantic comparison

The results discussed in this part give only a hint of what a deeper study of this kind would bring. Rather than drawing specific and complete conclusions, the main aim of this experimental section is to highlight some properties of the distances that would be worth considering when selecting a distance for a particular practical purpose.

We follow the technique described in [11] for analyzing the semantic similarities between dissimilarity measures between belief functions. N_s belief functions are randomly generated $\{\mathrm{Bel}^n\}_{n=1}^{N_s}$ (as described by **Algorithm 1**).

Input: X: Frame of discernment
Output: Bel: Belief function (under the form of a BPA, m)
Generate the power set of $X \rightarrow 2^X$;
Generate a random permutation of $2^X \rightarrow R(X)$;
Generate a integer between 1 and $N_{\max} \rightarrow k$;
foreach *First randomly generated k elements of* $R(X)$ **do**
| Generate a value within $[0, 1] \rightarrow m_k$;
end
Normalize the vector $\mathbf{m} = [m_1 \ldots m_k] \rightarrow \mathbf{m}^*$;
$m(A_k) = m_k$;

Algorithm 1: Random generation of a belief function.

We have generated five different types of BPAs: (1) simple support, *i.e.* $m(A) = \alpha$, $m(X) = 1 - \alpha$,(2) dichotomous, *i.e.* $m(A) = \alpha$, $m(\bar{A}) = \beta$, $m(X) = 1 - \alpha - \beta$, (3) complete, *i.e.* with $2^{|X|} - 1$ focal elements (4) with a fixed number of focal elements and (5) consonant, *i.e.* with nested focal elements. These types of BPAs have been used in the following to compare some behaviors of the distances in specific experiments.

The distances previously introduced in Section 3 and gathered in a set $\mathcal{D}$ are then computed for each of the N_s pairs (m^r, m^n), where m^r is a unique belief function of reference also randomly generated. Note that because the algorithms for random BPAs generation all involve a controlled number of focal elements (either 2, 3, 5 or $2^{|X|} - 1$) and that the masses are then uniformly assigned, the impact of the BPA of reference on the result provided in the following is very low.

The results presented in this section have been obtained for frames of discernment whose cardinality $|X|$ ranges from 2 to 8, for a number of replications N_s between 1000 and 10000 depending on the experiment, and for a set of distances of cardinality $|\mathcal{D}| = 15$. In $d_P^{(2)}$, the prior probability distribution has been assumed to be uniform over X so that $P(A, B) = \frac{N|A \cap B|}{|A|.|B|}$. In $d_{F(J)}^{(2)}$, F has been chosen as in [24]. Although in the original paper [46], for the function d_{PS}, the authors restricted the m_is to be simple support belief functions, we removed this restriction in our simulations.

Fig. 4 shows the scatter plots for each pair $\left(d_i(m^r, m^n), d_j(m^r, m^n)\right)$, $d_i, d_j \in \mathcal{D}$ for a cardinality of X of 8. The boxes on the diagonal of the scatter plot show the distributions of the measures. A straight line in the scatter plots means strong correlation (as for instance between d_{RS} and $\otimes_{Int}^d$, in accordance to their respective definitions) while a cloud of points means a weak correlation (as for instance between $\otimes_{Int}^d$ and $d_{Bet}^{(\infty)}$).

Due to the ordering of the distances according to their axiomatic properties, the correlation of the metrics and pseudometrics appear on the top of the figure. We observe a strong correlation between all the structural distances of the L_p family metrics and pseudometrics. No clear conclusion can be drawn whether the correlation is due to the family (L_p) or to the class (metric versus pseudometric), since we do not have enough examples of each class. We can simply notice a stronger correlation between the three nonstructural metrics whose weighting matrix are known similarity measures between the focal sets, say $d_J^{(2)}$, $d_{F(J)}^{(2)}$ and $d_D^{(2)}$ than with the nonstructural metric $d_{Inc}^{(2)}$.

The distances involving Dempster's conflict (and the conflict itself) $\otimes_{Int}^d$, d_{RS}, d_{PS} and d_{BP} are not correlated with the metric and pseudometric distances. Note that in this particular case of complete BPAs, the first member of the distance d_{PS} is always 0 and thus we only observe the second member $d_{PS(2)}$ in the scatter plots. For other kinds of BPAs, we would have observed steps corresponding to the integer values $d_{PS(1)}$.

[15] Note that here the values have been normalized relatively to $|\mathcal{F}_1 \cap \mathcal{F}_2| + 1$ which is the maximum value in this example, and not to $2^{|X|}$ so that the values are not too close to 0.

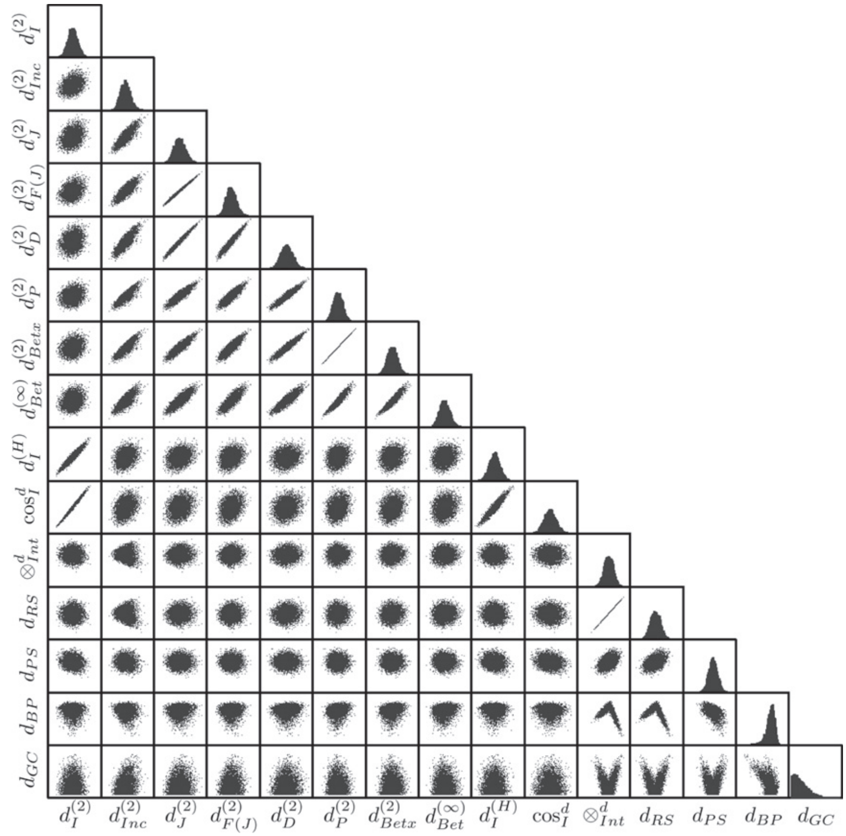


Fig. 4. Scatter plots for a frame of discernment of 8 elements and a random generation of complete BPAs.

The d_{GC} distance which qualifies as a semipseudometric seems to be noncorrelated with most of the distances apart from those involving Dempster's conflict, and would be perhaps a good candidate to be used in a 2D dimensional measure.

The scatter plots of Fig. 4 give an idea of the strength and direction of correlations between distances. In the following subsections, we will detail some particular features of these correlations on different kinds of BPAs. The comments below should then be read in correspondence to this scatter plots figure.

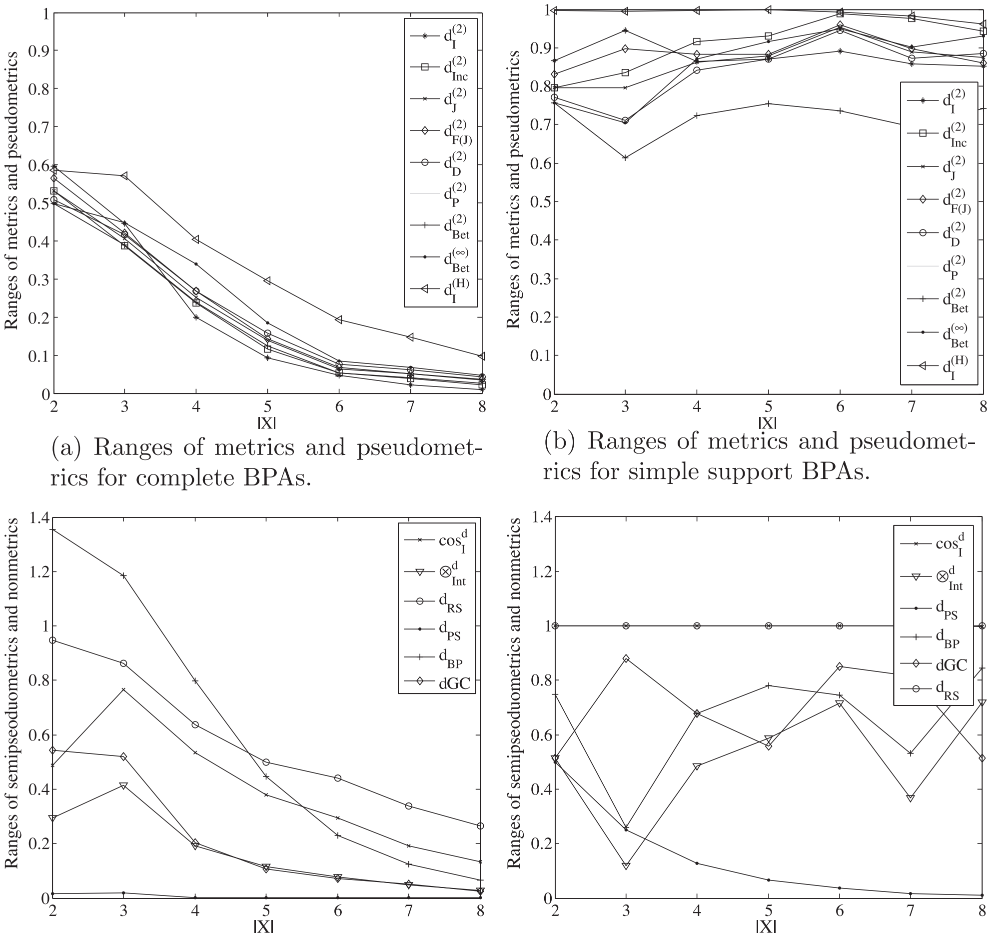
5.3. Distributions

The distributions for each distance appear on the diagonal of Fig. 4. Two main features of these distributions are discussed here: (1) the symmetry and (2) the range of values.

A nonsymmetric distribution means that higher (or lower) values of distances will be favored and that a median distance will not be in the middle of the interval range. The symmetrical properties of the distributions of distance values may have an impact in practice for instance if a threshold is used to select "close" belief functions to a reference one. This threshold should then be adjusted taking into account the symmetry of the distribution. Most of the distributions of distance values show a symmetric bell shape, except d_{BP} and d_{GC} whose distributions are nonsymmetric. The nonsymmetry of d_{GC} may be explained by the squared-root added to its definition, which was perhaps a too drastic and simple way to render it nonnegative and symmetric.

The range of the distance values may have an impact in practice especially when the distance is used in optimization or as a selection criterion. For instance, if the range of values is low, then the computations may be very sensitive to rounding errors when comparing distance values. Thus, a wide range is desirable. The range values (maximum minus minimum) are displayed in Fig. 5 for complete BPAs and simple support BPAs and for $|X|$ varying. It appears that for complete BPAs, the range decreases with the increase of $|X|$ even if the theoretical value is constant (see Table 6). This can be explained by the conjunction of the fact that the BPAs are complete and the constraint of the BPAs' coordinates which are positive and to sum up to 1. We also observe a particular behavior for d_{PS} whose range is very close to 0. Indeed, since the BPAs are complete, $d_{PS(1)}$ is always null. But $d_{PS(2)}$ is always below 1 and the distance has been normalized by $2^{|X|}$ which corresponds to the maximum value possibly reached by d_{PS}.

For simple support BPAs however, the range evolves differently. Indeed, because the number of focal elements is fixed to two, the range remains almost constant or at least neither increase nor decreases whatever the value of $|X|$. In practice, that



(a) Ranges of metrics and pseudometrics for complete BPAs.

(b) Ranges of metrics and pseudometrics for simple support BPAs.

(c) Ranges of semipseudometrics and nonmetrics for complete BPAs.

(d) Ranges of semipseudometrics and nonmetrics for simple support BPAs.

Fig. 5. Evolution of the ranges of the distances according to the cardinality of *X*. The statistics have been computed over 10000 replications.

Table 7
Correlation matrix of the distances corresponding to the scatter plot of Fig. 4, *i.e.* for $|X| = 8$ and $N_s = 6000$. The pair $(\otimes^d_{Int}, d^{(\infty)}_{Bet})$ proposed as a 2D measure of conflict in [41] has been highlighted in gray.

$d^{(2)}_I$	1.00														
$d^{(2)}_{Inc}$	0.30	1.00													
$d^{(2)}_J$	0.28	0.85	1.00												
$d^{(2)}_{F(J)}$	0.29	0.83	0.99	1.00											
$d^{(2)}_D$	0.22	0.85	0.98	0.96	1.00										
$d^{(2)}_P$	0.20	0.87	0.90	0.88	0.95	1.00									
$d^{(2)}_{Betx}$	0.20	0.87	0.90	0.88	0.95	1.00	1.00								
$d^{(\infty)}_{Bet}$	0.19	0.81	0.85	0.83	0.89	0.94	0.94	1.00							
$d^{(H)}_I$	0.95	0.29	0.28	0.28	0.22	0.20	0.20	0.19	1.00						
$\cos^d_I$	0.99	0.30	0.28	0.29	0.22	0.20	0.20	0.19	0.94	1.00					
$\otimes^d_{Int}$	−0.02	−0.09	0.06	0.06	0.06	0.07	0.07	0.07	−0.03	−0.02	1.00				
d_{RS}	−0.02	−0.09	0.06	0.06	0.06	0.07	0.07	0.07	−0.03	−0.02	1.00	1.00			
d_{PS}	−0.16	−0.10	−0.03	−0.02	−0.04	−0.04	−0.04	−0.04	−0.16	−0.17	0.42	0.42	1.00		
d_{BP}	−0.04	−0.13	0.05	0.04	0.09	0.15	0.15	0.15	−0.04	−0.04	−0.27	−0.27	−0.47	1.00	
d_{GC}	0.03	0.26	0.08	0.09	0.05	−0.01	−0.01	−0.01	0.03	0.03	0.07	0.07	0.09	−0.61	1.00
	$d^{(2)}_I$	$d^{(2)}_{Inc}$	$d^{(2)}_J$	$d^{(2)}_{F(J)}$	$d^{(2)}_D$	$d^{(2)}_P$	$d^{(2)}_{Betx}$	$d^{(\infty)}_{Bet}$	$d^{(H)}_I$	$\cos^d_I$	$\otimes^d_{Int}$	d_{RS}	d_{PS}	d_{BP}	d_{GC}

means that in an application in which the number of focal elements is intended to grow (like in a long sequential combination process), then we must be conscious that the algorithm may be more sensitive to rounding errors as the cardinality of *X* is high. Perhaps, a kind of normalization relatively to the number of focal elements may be considered.

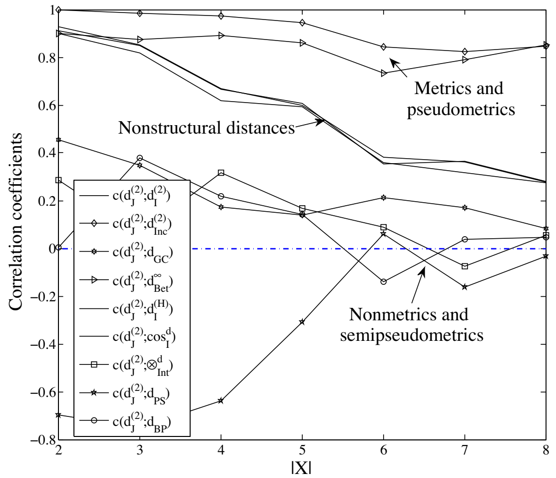


Fig. 6. Evolution of the correlation coefficient between the metric $d_J^{(2)}$ and a representative subset of other distances according to the cardinality of $|X|$.

Besides the range, higher-order statistics such as the standard deviation, the skewness and the kurtosis are interesting: The standard deviation is a measure of spread and according to equivalent arguments in favour of a wide range, a large spread may be desirable; the skewness is a measure of symmetry and could be studied more deeply for applications where the symmetry of distance values is involved; and the kurtosis is a measure of "peakedness". For instance, in a preliminary study we observed the effect of the F function on $d_J^{(2)}$ according to these statistics: $d_{F(J)}^{(2)}$ is more spread $(0.0075 > 0.0071)$, less symmetric $(0.252 < 0.285)$ and less peaked $(0.484 < 0.516)$ than $d_J^{(2)}$. The impact of the F function over **Jac** can also be observed in box $(1, 1)$ in Fig. 8.

5.4. Correlation

For each pair of measures in $\mathcal{D}$, we computed a Pearson coefficient, and built corresponding matrix whose elements are:

$$c(d_i, d_j) = \frac{\sum_{n=1}^{N}(d_i^n - \bar{d}_i)(d_j^n - \bar{d}_j)}{\sqrt{\sum_{n=1}^{N}(d_i^n - \bar{d}_i)^2}\sqrt{\sum_{n=1}^{N}(d_j^n - \bar{d}_j)^2}} \quad (69)$$

with $\bar{d}_i = \frac{1}{N}\sum_{n=1}^{N} d_i$. If $c(d_i, d_j) = 0$, the two distances are uncorrelated while the distances are all the more correlated when $c(d_i, d_j)$ is close to 1 (or -1). A negative value for $c(d_i, d_j)$ means that d_i and d_j are in dissimilarity/similarity correspondence, while a positive value means that they are either in dissimilarity/dissimilarity correspondence or in similarity/similarity correspondence. As an illustration, a characteristic of nonmetric distances is that their correlation coefficient relatively to a known dissimilarity measure is either positive or negative. This behavior can be seen in Fig. 6, relatively to the metric $d_J^{(2)}$.

Table 7 lists the correlation coefficients corresponding to the scatter plots of Fig. 4.

A low correlation (close to 0) between two measures means that they quantify two distinct (and possibly complementary) aspects of the distance between two belief functions while a high correlation means that they are redundant. Hence, weakly correlated pairs of distances could be good candidates for two dimensional measures. The pair $(\otimes_{Int}^d, d_{Bet}^{(\infty)})$ proposed in [41] has been highlighted in gray in the table. It has a quite low correlation and is thus justified for a 2D measure, according this criterion (see also Fig. 9). However, alternative pairs could be interesting candidates such as $(d_{Bet}^{(\infty)}, d_{GC})$, $(d_{Bet}^{(2)}, d_{PS})$, $(d_P^{(2)}, d_{PS})$ and $(d_{Bet}^{(\infty)}, d_{BP})$ as their correlation coefficients are very close to 0.

Fig. 6 shows the evolution of the correlation coefficient between a metric of reference $d_J^{(2)}$ and all the others while $|X|$ increases. We chose $d_J^{(2)}$ as a representative distance of the class of structural metrics, *i.e.* distances satisfying the higher number of properties. Any other distance in this class could have been chosen instead, such as $d_{Inc}^{(2)}$. Due to the strong correlation of the distances in that class, the results obtained with $d_{Inc}^{(2)}$ for instance were very similar to the one presented here in Figs. 6 and 8. For a clearer visualization, we selected a subset of distances representative of their respective classes. In particular, we dropped one of the two obviously strongly correlated pairs of distances, say $(d_J^{(2)}, d_{F(J)}^{(2)})$, $(d_{Betx}^{(2)}, d_P^{(2)})$, $(\otimes_{Int}^d, d_{RS})$. Note that we observed similar results with the metric $d_{Inc}^{(2)}$.

In general, the correlation decreases as $|X|$ increases. The nonmetric distances point out their nature since their correlation coefficient is either positive or negative meaning that they are either in dissimilarity/dissimilarity correspondence or in similarity/dissimilarity with a known dissimilarity measure (here a full metric). We observe that the evolution of the correlation coefficient is very similar for all the nonstructural distances whatever their metric properties. Also, the metrics and pseudometrics have similar behavior.

Fig. 7 shows the evolution of the correlation coefficient of the pair $(d_{Bet}^{(\infty)}, \otimes_{Int}^{d})$ defined as a 2D measure in [41], for four different kinds of BPAs: Simple support, consonant, dichotomous and complete.

It appears that for consonant BPAs, both measures are highly correlated although the correlation decrease slightly as $|X|$ increases. In this case, the 2D measure may be not so much informative. However, the correlation is relatively low for complete BPAs and even lower as $|X|$ is high, which makes the 2D measure very relevant in that case. Finally, nothing can be said in the case of simple support and dichotomous BPAs although the correlation seems to oscillate around 0.5 in the latter case.

Fig. 8 shows the superposition of the scatter plots for several cardinalities of X, for a structural metric chosen as representative of its class $(d_J^{(2)})$ against three types of distances: Other L_2 metrics (first column), nonmetric distances or a Fidelity measure (second column) and L_2 pseudometrics (third column).

As already observed in Fig. 5, the ranges of values for all the measures decrease as $|X|$ increases. This is due to the particular kind of belief functions (complete BPAs) used in the simulations.

In light of the combined results of Figs. 6, 7 and 8, several remarkable cases need to be discussed:

- $c(d_J^{(2)}, d_{Inc}^{(2)})$ – $\diamond$ in Fig. 6 and box (4, 1) in Fig. 8: While the correlation between these two structural metric distances is very high for $|X| = 2$, it decreases sightly as $|X|$ increases. Moreover, the slope a increases with $|X|$: It moves from a $y = x$ relation for $|X| = 2$ to a $y = ax$ relation with $a > 1$, and thus the one-to-many relation increases. In other words, the range of $d_J^{(2)}$ decreases more than that of $d_{Inc}^{(2)}$. Hence, for high cardinalities a single value for $d_J^{(2)}$ corresponds to many distinct values for $d_{Inc}^{(2)}$.
- $c(d_J^{(2)}, d_{Bet}^{(\infty)})$ – $\triangleright$ in Fig. 6 and box (3, 3) in Fig. 8: We observe here the difference between structural metric $d_J^{(2)}$ and structural pseudometric and see that the latter yields several values when the former gives only one. The same behavior occurs for the other pseudometrics.
- $c(d_J^{(2)}, \otimes_{Int}^{d})$ – $\square$ in Fig. 6 and box (2, 2) in Fig. 8: The nonmetric characteristic of $\otimes_{Int}^{d}$ mentioned in Section 3.3 is illustrated here as the correlation is either positive (dissimilarity) or negative (similarity). The same behavior can be observed for d_{PS} (and also for d_{BP} although not shown here).
- $c(d_J^{(2)}, d_I^{(H)})$ – — in Fig. 6 and box (4, 2) in Fig. 8: This illustrates the comparative behavior between two metrics of two distinct families, namely L_2 and Fidelity respectively. The correlation between these two distances decreases a lot as $|X|$ increases. Moreover, the values of $d_I^{(H)}$ are clearly distinct for distinct sizes of X.
- $c(d_J^{(2)}, d_I^{(2)})$ – — in Fig. 6 and box (3, 1) in Fig. 8: This illustrates the behaviour of two metrics of the same family L_2, one structural $d_J^{(2)}$ and one nonstructural $d_I^{(2)}$. The correlation between these two distances is very similar to the correlation between two other L_2 distances (for instance $d_J^{(2)}$ and $d_D^{(2)}$) while a different behavior is observed between two distances of distinct families $d_J^{(2)}$ of L_2 and $d_I^{(H)}$ of Fidelity. This may be a clue for a deeper analysis of the impact of the families of distances with respect to their metric or structural properties.

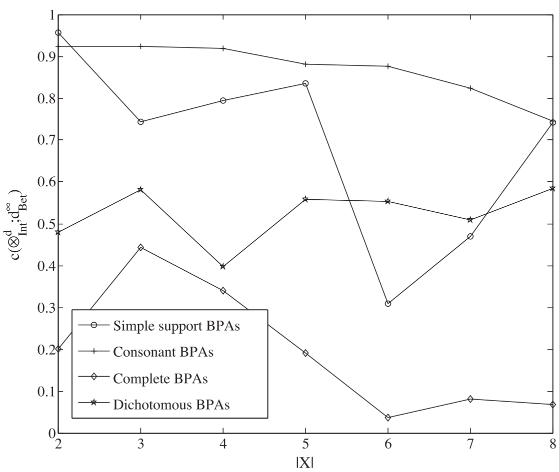


Fig. 7. Evolution of the correlation coefficient of the two dimensional measure $(d_{Bet}^{(\infty)}, \otimes_{Int}^{d})$ according to the cardinality of $|X|$ for several kinds of BPAs.

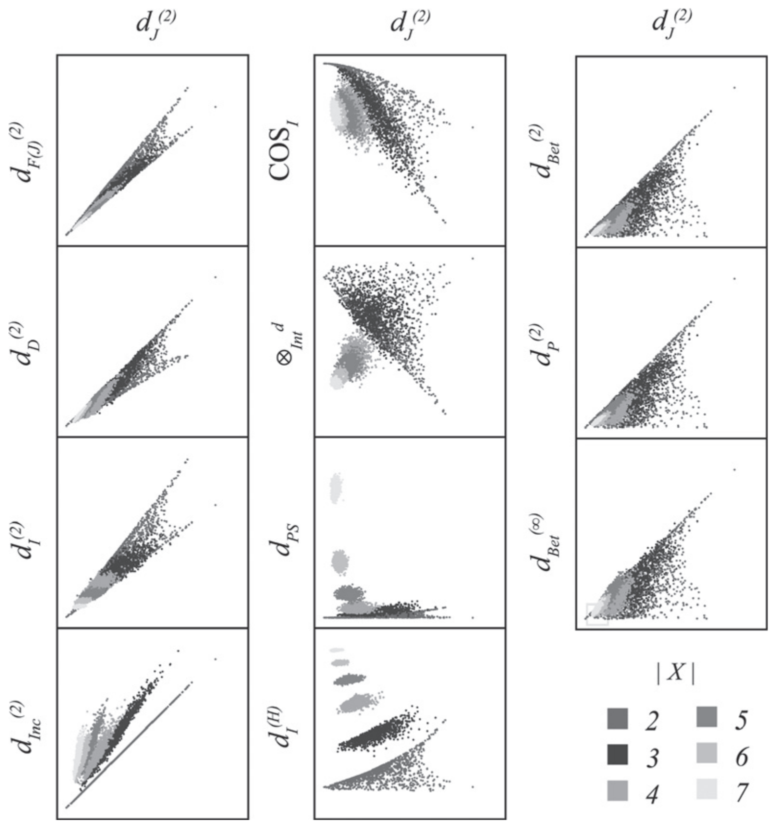


Fig. 8. Superposition of the scatter plots for several cardinalities of X (from 2 to 7) with 1000 replications of complete BPAs. Three groups of distances are compared to a structural metric chosen as reference, $d_J^{(2)}$.

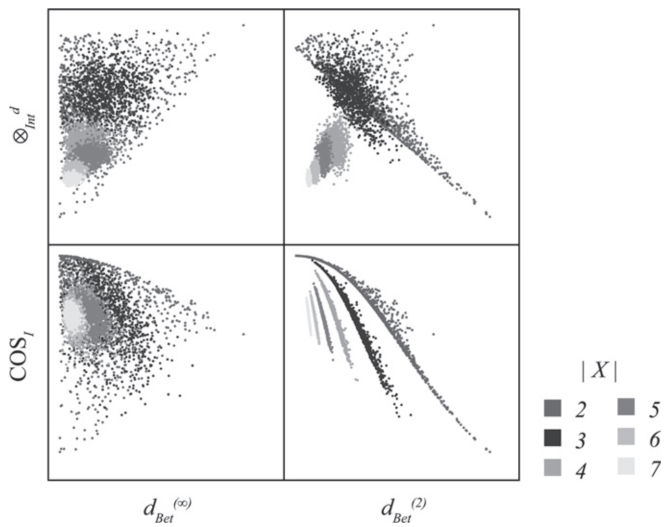


Fig. 9. Superposition of the scatter plots for several cardinalities of X (from 2 to 7) between four distances: Two pseudometrics of the same family, one similarity measure and Dempster's conflict.

Fig. 9 shows the scatter plots between several classes of measures: Two pseudometrics belonging to the same family L_p, $d_{Bet}^{(\infty)}$ (L_∞) and $d_{Betx}^{(2)}$ (L_2), Dempster's conflict $\otimes_{Int}^d$ and the similarity measure $\cos_I^s = 1 - \cos_I^d$. We first observe that although the two pseudometrics are based on the same weighting matrix $\mathbf{Bet}'\mathbf{Bet}$, these behave differently when compared to the two other distances. This highlights the impact of the p coefficient within the L_p family. Maybe the latter could be the subject of further investigations.

The correlation $c(d_{Bet}^{(2)}, \cos_I^s)$ is remarkable as it remains high whatever the cardinality of X while it is very distinct from one cardinality to another.

The couple $(d_{Bet}^{(\infty)}, \otimes_{Int}^d)$ corresponds to Liu's 2D index and should be completed by the curve with diamonds in Fig. 7 corresponding to complete BPAs. In both graphics, we see that the correlation remains very low as $|X|$ increases. However, we also observe that the space in the scatter plot is only half-covered (upper-right triangle) meaning that a kind of correlation exists between these two measures. The Pearson coefficient is only one measure of similarity among others, and behaviors such as this one suggest studying other measures of similarity between distances.

5.5. Additive trees

Based on the correlation coefficient values, we built the additive trees (dendrograms) for all six cardinalities of X (Fig. 10). The dendrograms are built using the correlation coefficient given by Eq. (69) as a similarity measure. For the construction and interpretation of the trees the method presented in Sattath and Tversky's classical paper [51] is used. The dissimilarity

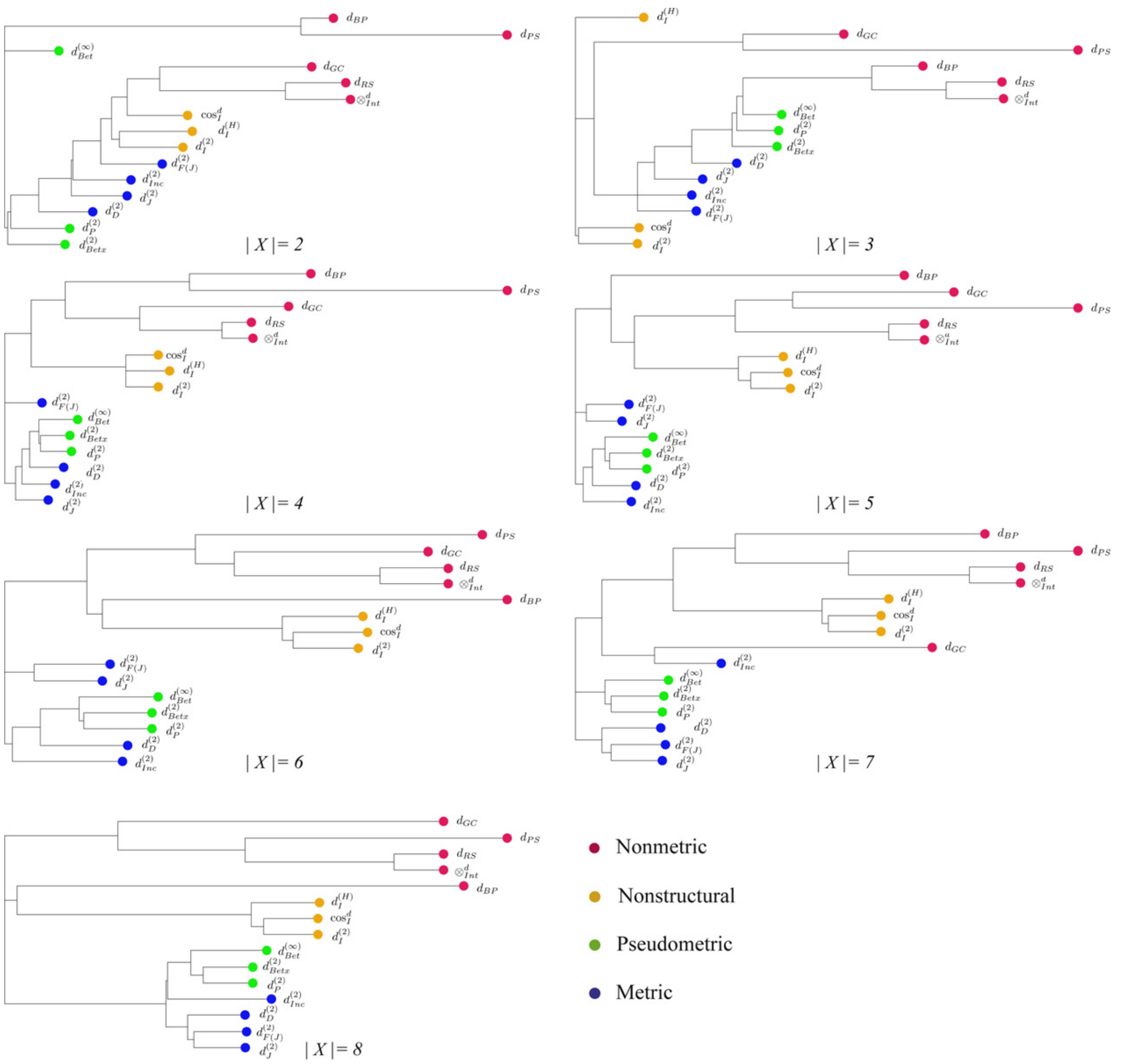


Fig. 10. Additive trees computed on a Pearson correlation matrix between 10000 replications based on Algorithm 1 generating complete BPAs over a frame of discernment whose cardinality varies from 2 to 8.

of two objects represented on these additive trees is simply given by the length of the path joining them, the longer the path, the more dissimilar are the objects under study.

Referring to the classes identified in Table 5, four groups of distances are considered: Nonmetrics (red circles), pseudometrics (green circles), metrics (blue circles) and nonstructural distances (yellow circles). These four groups clearly appear in the trees although the tree structure is slightly sensitive to the size of the frame of discernment. Rather than the family (L_p, inner product or Fidelity) it appears that the axiomatic properties play a major role in the similarity of the distances. Moreover, the nonstructural property seems even stronger than the metric properties since the three nonstructural distances remain clearly regrouped as $|X|$ varies while one is a dissimilarity ($\cos_I^d$), one is a L_2 metric ($d_I^{(2)}$) and one is a Fidelity metric ($d_I^{(H)}$).

Metrics and pseudometrics are very close to each other and form two very compact groups. Nonmetric distances are not so well assembled in a compact group but seem to behave very differently from the others. This is very clear for d_{PS} while d_{GC} is sometimes closer to the other kinds of distances. We must outline that the nonmetric distances have another feature in common apart from their nonmetric properties, which is that they all involve Dempster's conflict. A deeper analysis including other nonmetric measures not based on $\otimes_{Int}^d$ or metric measures based on $\otimes_{Int}^d$ would be required to discriminate the influence of $\otimes_{Int}^d$ relatively to the metric properties.

The results of the experimentations presented here are not exhaustive and a lot of work remains to be done. However, these preliminary results highlighted some remarkable experimentally observed features that we have been able to relate to the theoretical properties. In practice, these results suggest that the choice of a measure should be guided by (1) the structural properties, a structural distance being in general more desirable for the quantification of the distance between two belief functions, (2) the metric properties since the definiteness may be also desirable in some cases. Regardless these "technical" properties, semantic properties would also be worth some consideration as the meaning of the distance may be of prime importance in some applications. For instance, d_{GC} quantifies a difference in information contents which is very different from $d_{Bet}^{(p)}$ which quantifies the difference in decision abilities. It is not clear however how important is the family the distances belong to, although we have been able to notice some impact of the p coefficient in the L_p family for instance.

6. Conclusions and future work

In this paper we have presented, together with a synthesis of their most important properties, a survey of the main dissimilarity measures defined so far using the mathematical framework of evidence theory. The principal results of this study are briefly recapitulated here, followed by a short discussion on future work.

We have first outlined the existence of a formal link between the existing distances defined so far in the mathematical framework of belief functions and the theory of inner products. This helped us propose a grouping of the dissimilarity measures into five families, *i.e.* Composite, Minkowski, Inner product, Fidelity and Information-based, families for which we have provided general formulations. Building upon these generalizations we have subsequently defined more than 40 new distances. Then, we have proposed a classification based on the one hand (a) on the study of the metric properties of the surveyed distance measurements allowing us to distinguish four groups (metrics, pseudometrics, semipseudometrics and nonmetric measures) and on the other hand (b) on structural properties allowing us the distinguish two groups of measurements (structural and nonstructural). This classification has been used later in the paper to interpret the results of simulations aimed at the experimental comparison of the surveyed distance measurements.

We have also provided a general expression for two-dimensional indexes expressed as couples composed of a cosine and a dissimilarity and proposed an alternative formulation to Dempster's conflict based on a cosine measure. Using a toy example about the monitoring of the convergence to a categorical belief function, we have illustrated the behavior of the different families of distance measurements.

Experimental comparisons based on the simulation of randomly assigned belief masses for different frame of discernment cardinalities allowed us to highlight the practical effects of the theoretical properties outlined in the present work, in particular (a) that of the non metric property of Dempster's conflict, as well as (b) the high correlation of measurements obtained within the same formal family. We have thus been able to relate experimental observations to some of the theoretical properties of the measures studied, admittedly an unexpected finding for the authors.

This paper can be seen as a first attempt to synthesize the theoretical properties of the existing distance measurements in the mathematical framework of evidence theory. Much work remains to be done, either on the theoretical or on the practical front of this emerging specialty. In future work we will refine the theoretical underpinnings outlined here, and pursue the study of the new distance measurements discovered in the present study.

An important endeavor is to develop practical applications using the newly discovered measurements and to study their behavior under different contexts of use. In the short term, future experimental work will include the development and comparison of clustering and information retrieval methods.

Acknowledgments

We gratefully thank Dr. Mihai Florea for many useful discussions and for his support for some of the programming. We would also like to thank the anonymous reviewers for their relevant and constructive comments which have helped us improve the paper considerably.

References

[1] J. Abellán, M. Gómez, Measures of divergence on credal sets, Fuzzy Sets and Systems 157 (2006) 1514–1531.
[2] M. Basseville, Distance measures for signal processing and pattern recognition, European Journal Signal Processing 18 (1989) 349–369.
[3] M. Bauer, Approximation algorithms and decision making in the Dempster–Shafer theory of evidence – an empirical study, International Journal of Approximate Reasoning 17 (1997) 217–237.
[4] S. Ben-Hariz, Z. Elouedi, K. Mellouli, Clustering approach using belief function theory, in: J. Euzenat, J. Domingue (Eds.), Artificial Intelligence: Methodology, Systems, and Applications, Lecture Notes in Computer Science, vol. 4183, 2006, pp. 162–171.
[5] M.J. Beynon, The role of the DS/AHP in identifying inter-group alliances and majority rule within group decision making, Group Decision and Negotiation 15 (2006) 21–42.
[6] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, Bulletin of the Calcutta Mathematical Society 35 (1943) 99–109.
[7] S. Blackman, R. Popoli, Design and Analysis of Modern Tracking Systems, Artech House, 1999.
[8] I. Bloch, On fuzzy distances and their use in image processing under imprecision, Pattern Recognition 32 (1999) 1873–1895.
[9] M. Bouchard, A.-L. Jousselme, P.-E. Doré, A proof for the positive definiteness of the Jaccard index matrix, International Journal of Approximate Reasoning, submitted for publication.
[10] L.M. de Campos, M.T. Lamata, S. Moral, Distances between fuzzy measures through associated probabilities: some applications, Fuzzy Sets and Systems 35 (1990) 57–68.
[11] S.-H. Cha, Comprehensive survey on distance/similarity measures between probability density functions, International Journal of Mathematical Models and Methods in Applied Sciences 1 (2007) 300–307.
[12] L.-Z. Chen, W.-K. Shi, Y. Deng, Z.-F. Zhu, A new fusion approach based on distance of evidences, Journal of Zhejiang University Science 6A (2005) 476–482.
[13] F. Cuzzolin, Two new Bayesian approximations of belief functions based on convex geometry, IEEE Transactions on Systems, Man, and Cybernetics part B 37 (2007) 993–1008.
[14] F. Cuzzolin, A geometric approach to the theory of evidence, IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews 38 (2008) 522–534.
[15] F. Cuzzolin, Geometry of Dempster's rule of combination, IEEE Transactions on Systems, Man, and Cybernetics – Part B: Cybernetics 34 (2008) 961–977.
[16] F. Cuzzolin, Consistent approximations of belief functions, in: 6th International Symposium on Imprecise Probability: Theories and Applications. Durham, United Kingdom, 2009.
[17] A. Dempster, Upper and lower probabilities induced by multivalued mapping, The Annals of Mathematical Statistics 38 (1967) 325–339.
[18] Y. Deng, D. Wang, Q. Li, An improved combination rule in fault diagnosis based on Dempster–Shafer theory, in: Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, 2008.
[19] T. Denœux, A neural network classifier based on Dempster–Shafer theory, IEEE Transactions on Systems, Man and Cybernetics 30 (2000) 131–150.
[20] T. Denœux, Inner and outer approximation of belief structures using a hierarchical clustering approach, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 9 (2001) 437–460.
[21] T. Denœux, Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence, Artificial Intelligence 172 (2008) 234–264.
[22] T. Denœux, A. Ben Yaghlane, Approximating the combination of belief functions using the fast Mœbius transform in a coarsened frame, International Journal of Approximate Reasoning 31 (2002) 77–101.
[23] T. Denœux, M. Masson, EVCLUS: evidential clustering of proximity data, IEEE Transactions on Systems, Man and Cybernetics Part B 34 (2004) 95–109.
[24] J. Diaz, M. Rifqi, B. Bouchon-Meunier, A similarity measure between basic belief assignments, in: Proceedings of the 9th International Conference Information Fusion, Firenze, Italy, 2006.
[25] D. Dubois, H. Prade, Consonant approximations of belief functions, International Journal of Approximate Reasoning 4 (1990) 419–449.
[26] D. Fixsen, R.P.S. Mahler, The modified Dempster–Shafer approach to classification, IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans 27 (1997) 96–104.
[27] M.C. Florea, E. Bossé, Crisis management using Dempster Shafer theory: using dissimilarity measures to characterize sources' reliability, in: C3I in Crisis, Emergency and Consequence Management, RTO-MP-IST-086. Bucharest, Romania, 2009.
[28] M.C. Florea, E. Bossé, A.-L. Jousselme, Metrics, Distances and dissimilarity measures within Dempster–Shafer theory to characterize sources' reliability, in: Cognitive Systems with Interactive Sensors (COGIS'09), 2009.
[29] M.C. Florea, A.-L. Jousselme, E. Bossé, Dynamic estimation of evidence discounting rates based on information credibility, RAIRO Operations Research 44 (2011) 285–306.
[30] A.D. Gordon, Constructing dissimilarity measures, Journal of Classification 7 (1990) 257–269.
[31] X. Guan, X. Yi, Y. He, Research on conflicting evidences combination strategies, in: Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, 2008.
[32] D. Harmanec, Faithful approximations of belief functions, in: K.B. Laskey, H. Prade (Eds.), Uncertainty in Artificial Intelligence (UAI99) 15, Stockholm, Sweeden, 1999.
[33] E. Hellinger, Neaue begründung der theorie der quadratischen formen von unendlichen vielen veränderlichen, Journal für die Reine und Aangewandte Mathematik 136 (1909) 210–271.
[34] I. Jenhani, S. Benferhat, Z. Elouedi, Properties analysis of inconsistency-based possibility similarity measures, in: J.V.L. Magdalena, M. Ojeda-Aciego, (Eds.), Proceedings of IPMU'08, Torremolinos (Málaga), 2008, pp. 173–180.
[35] A.-L. Jousselme, D. Grenier, E. Bossé, A new distance between two bodies of evidence, Information Fusion 2 (2001) 91–101.
[36] A.-L. Jousselme, P. Maupin, On some properties of distances in evidence theory, in: Workshop on the Theory of Belief Functions, Brest, France, 2010.
[37] D.G. Kendall, Foundations of a theory of random sets, in: E. Harding, D. Kendall (Eds.), Stochastic Geometry, Wiley, New-York, 1974, pp. 322–376.
[38] J. Klein, O. Colot, Automatic discounting rate computation using a dissent criterion, in: Proceedings of the first Workshop on the Theory of Belief Functions, Brest, France, 2010.
[39] G.J. Klir, Uncertainty and Information – Foundations of Generalized Information Theory, Wiley-Interscience, 2005.
[40] G.J. Klir, M.J. Wierman, Uncertainty-based information, Studies in Fuzziness and Soft Computing, vol. 15, second ed., Physica-Verlag, Heidelberg, New York, 1999.
[41] W. Liu, Analyzing the degree of conflict among belief functions, Artificial Intelligence 170 (2006) 909–924.
[42] R.P.S. Mahler, Combining ambiguous evidence with respect to ambiguous a priori knowledge. I: boolean logic, IEEE Transactions on Systems, Man and Cybernetics 26 (1996) 27–41.
[43] A. Martin, A.-L. Jousselme, C. Osswald, Conflict measure for the discounting operation on belief functions, in: Proceedings of the 11th Annual Conference on Information Fusion, Cologne, Germany, 2008.
[44] G. Mathéron, Random Sets and Integral Geometry, Wiley, New-York, 1975.
[45] D. Mercier, B. Quost, T. Denœux, Refined modeling of sensor reliability in the belief function framework using contextual discounting, Information Fusion 9 (2008) 246–258.
[46] W.L. Perry, H.E. Stephanou, Belief function divergence as a classifier, in: Proceedings of the 1991 IEEE International Symposium on Intelligent Control, Arlington, VA, USA, 1991, pp. 280–285.
[47] S. Petit-Renaud, T. Denœux, Nonparametric regression analysis of uncertain and imprecise data using belief functions, International Journal of Approximate Reasoning 35 (2004) 1–28.

[48] F. Pichon, T. Denœux, The unnormalized Dempster's rule of combination: a new justification from the least commitment principle and some extensions, Journal of Automated Reasoning 45 (2010) 61–87.
[49] B. Quost, T. Denoeux, M.-H. Masson, Pairwise classifier combination in the transferable belief model, in: Proceedings of the 7th International Conference on Information Fusion, Philadelphia, USA, 2005.
[50] B. Ristic, P. Smets, The TBM global distance measure for the association of uncertain combat ID declarations, Information Fusion 7 (2006) 276–284.
[51] S. Sattah, A. Tversky, Additive similarity trees, Psychometrika 42 (1977) 319–345.
[52] J. Schubert, Clustering decomposed belief functions using generalized weights of conflict, International Journal of Approximate Reasoning 48 (2008) 466–480.
[53] P. Smets, The application of the matrix calculus to belief functions, International Journal of Approximate Reasoning 31 (2002) 1–30.
[54] P. Smets, Analyzing the combination of conflicting belief functions, Information Fusion 8 (2007) 387–412.
[55] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191–234.
[56] H.E. Stephanou, S.-Y. Lu, Measuring consensus effectiveness by generalized entropy, IEEE Transaction on Pattern Analysis and Machine Intelligence 10 (1988) 544–554.
[57] B. Tessem, Approximations for efficient computation in the theory of evidence, Artificial Intelligence 61 (1993) 315–329.
[58] A. Tversky, Features of similarity, Psychological Review 84 (1977) 327–352.
[59] C. Wen, Y. Wang, X. Xu, Fuzzy information fusion algorithm of fault diagnosis based on similarity measure of evidence, in: Advances in Neural Networks, Lecture Notes in Computer Science vol. 5264, Springer, Berlin/ Heidelberg (2008) 506–515.
[60] Q. Ye, X. Wu, Z. Chen, An approach for evidence clustering using generalized distance, Journal of Electronics 26 (2009) 18–23.
[61] L.M. Zouhal, T. Denœux, An evidence-theoretic *k*-NN rule with parameter optimization, IEEE Transactions on Systems, Man and Cybernetics – Part C 28 (1998) 263–271.