

Evidential Matrix Metrics as Distances Between Meta-Data Dependent Bodies of Evidence

Mehena Loudahi, John Klein, Jean-Marc Vannobel, Olivier Colot

► **To cite this version:**

Mehena Loudahi, John Klein, Jean-Marc Vannobel, Olivier Colot. Evidential Matrix Metrics as Distances Between Meta-Data Dependent Bodies of Evidence. IEEE transactions on systems, man, and cybernetics, Institute of Electrical and Electronics Engineers (IEEE), 2016, 46 (1), pp.109 - 122. <10.1109/TCYB.2015.2395877>. <hal-01115637>

HAL Id: hal-01115637

<https://hal.archives-ouvertes.fr/hal-01115637>

Submitted on 11 Feb 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Evidential matrix metrics as distances between meta-data dependent bodies of evidence

Mehena Loudahi, John Klein, Jean-Marc Vannobel and Olivier Colot ^{*†}

Abstract

As part of the theory of belief functions, we address the problem of appraising the similarity between bodies of evidence in a relevant way using metrics. Such metrics are called evidential distances and must be computed from mathematical objects depicting the information inside bodies of evidence. Specialization matrices are such objects and, therefore, an evidential distance can be obtained by computing the norm of the difference of these matrices. Any matrix norm can be thus used to define a full metric.

In this article, we show that other matrices can be used to obtain new evidential distances. These are the α -specialization and α -generalization matrices and are closely related to the α -junctive combination rules. We prove that any L^1 norm-based distance thus defined is consistent with its corresponding α -junction. If $\alpha > 0$, these distances have in addition relevant variations induced by the poset structure of the belief function domain. Furthermore, α -junctions are meta-data dependent combination rules. The meta-data involved in α -junctions deals with the truthfulness of information sources. Consequently, the behavior of such evidential distances is analyzed in situations involving uncertain or partial meta-knowledge about information source truthfulness.

Index terms— Theory of belief functions, evidence theory, distance, combination rules, information source meta-data.keywords

1 Introduction

The theory of belief functions [9, 26] is a framework for reasoning under uncertainty and imprecision. It encompasses probability theory, possibility theory and Cantor's set theory. Approaches relying on the theory of belief functions are often called *evidential* because collected data are seen as pieces of evidence allowing the definition of some set functions describing the state of knowledge

^{*}M. Loudahi, J. Klein, J-M. Vannobel and O. Colot are with Lille1 University and CRISAL UMR CNRS 9189, Lille, France (e-mail: firstname.lastname@univ-lille1.fr).

[†]Manuscript accepted for publication in IEEE Transactions on Cybernetics on january 2015.

of a given information source concerning the problem at stake. The pieces of evidence collected by each source individually are called bodies of evidence.

In particular, in case of multiple information sources, it is sometimes necessary to determine if the contents of different bodies of evidence are similar or not. A distance between bodies of evidence can achieve this goal and such distances are called evidential distances.

Defining a meaningful evidential distance is a challenging task as dissimilarity has its roots in two main aspects: uncertainty and imprecision. For instance, the statement *John is 30 years old* is different in terms of uncertainty from the statement *John is 30 years old with probability 0.6*. The statement *John is 30 years old* is different in terms of imprecision from the statement *John is in his thirties*. It is obvious that a positive distance value is justified in both of these comparisons but it is far less obvious to determine if one of them should prevail or not. When introducing evidential distances, authors do not always justify their approach with respect to uncertainty or imprecision although their influences are latent. To some extent, defining evidential distances is thus an over-constrained problem with many possible heterogeneous solutions depending on a targeted purpose. Such purposes can be mathematical properties or applications. This accounts for the vast literature dealing with evidential distances.

The most popular evidential distance is Josselme distance [16]. Belief functions are set functions and their domain has a poset structure. Josselme distance is a full metric and takes explicitly into account this poset structure. Prior to Josselme *et al.* approach, people used to resort to pseudo-metrics [30, 22, 34, 32, 12] and [1]. One full metric [15, 19] was introduced before Josselme distance and used inside an optimization algorithm. Discussing its relevance for assessing belief function closeness was not in the scope of that paper. More recently, Diaz *et al.* introduced in [10] several metrics in the same vein as Josselme distance but with different appraisals of structural aspects.

Most of the time, evidential distances are application-driven. For instance, Zouhal and Denœux [34] use a nearest neighbor classifier where neighbor testimonies are seen as bodies of evidence. The parameters involved in their evidential approach are optimized by minimizing a dissimilarity between these bodies. Fixen and Mahler [12] introduce a pseudo-metric to evaluate the performances of their classifier whose outputs are bodies of evidence. In [11] and [31], sensor data are represented as bodies of evidence and the reliability of the sensor is assessed using distances. A common issue when working in the framework of belief functions consists in approximating a body of evidence in order to obtain another one with desired properties or in order to reduce computation load. A convenient way to perform such approximations is to resort to distance minimization as proposed in [15, 1, 33],[32, 5, 7] or [8].

past decade.

Josselme and Maupin [17] analyzed and surveyed the evidential distance literature and clarified the advantages and limitations of evidential distances. Based on their comments, we formalized in [21] a number of desirable mathematical properties for evidential distances and proposed specific distances possessing

these properties. These distances are obtained using the norm of the difference between matrices encoding bodies of evidence. In particular, one of these new distances has similar properties as Jousselme distance while exhibiting a consistent behavior with respect to the conjunctive rule of combination. In this article, we continue our effort toward the definition of meaningful evidential distances by investigating the properties of distances derived using other matrices than those considered in [21].

After providing the necessary definitions and background related to the theory of belief functions in section 2, extended families of matrix-based evidential distances are presented in section 3. These families are infinite and are defined using evidential matrices that are linked with combination rules known as α -junctions. In section 4, the consistency of the proposed distances with respect to α -junctions is investigated. Let \mathcal{B}_0 , \mathcal{B}_1 and \mathcal{B}_2 denote three bodies of evidence. In this paper, we say that an evidential distance is consistent with a combination rule if the distance between \mathcal{B}_1 and \mathcal{B}_2 is greater or equal to the distance between $\{\mathcal{B}_1$ combined with $\mathcal{B}_0\}$ and $\{\mathcal{B}_2$ combined with $\mathcal{B}_0\}$. In section 5, we discuss the ability of the proposed distances to take the poset structure of belief functions domain into account. The importance of this aspect is stressed in [17]. Finally, in section 6, evidential distances are also studied under a new perspective: the influence of meta-information. Meta-information is made of contextual data on the information sources. Various scenarios on the truthfulness of sources are considered.

2 Belief function framework: notations and definitions

In this section, some mathematical notations for classical belief function concepts are given. The reader is expected to be familiar with belief function basics and consequently some definitions are not detailed. More material on belief functions basics is found for instance in [3]. Special care is given to a reminder on α -junctions and matrix calculus as part of the theory of belief functions.

2.1 Belief function basics

Suppose $\Omega = \{\omega_i\}_{i=1}^n$ is the finite domain of a given parameter θ , *i.e.* the set of possible and unknown values ω_i of θ . Suppose also S_i is a source of information that collected a body of evidence \mathcal{B}_i regarding the actual value of θ . The set function representing all pieces of evidence in \mathcal{B}_i is called a **mass function** and is denoted by m_i . The power set 2^Ω is the set of all subsets of Ω and it is the domain of mass functions. For any $A \in 2^\Omega$, the **cardinality** of this set is denoted by $|A|$ and we thus have $|\Omega| = n$. The cardinality of 2^Ω is denoted by $N = 2^n$. Mass functions have $[0, 1]$ as co-domain and they sum to one: $\sum_{A \in 2^\Omega} m_i(A) = 1$. A **focal element** of a mass function m_i is a set $A \subseteq \Omega$ such that $m_i(A) > 0$, meaning that \mathcal{B}_i contains evidence supporting the event

$\{\theta \in A\}$. A mass function having only one focal element A is called a **categorical mass function** and is denoted by m_A . The categorical mass function m_Ω is called the **vacuous mass function** because it carries no information. As explained in [6], a **simple mass function** is the convex combination of m_Ω with a categorical mass function m_A with $A \neq \Omega$.

In the theory of belief functions, several alternative set functions are commonly used for evidence representation. In this paper, only the **plausibility** function pl_i will be used. This set function is the conjugate of the inverse Möbius of the mass function m_i . The plausibility function pl_i is in one-to-one correspondence with the mass function m_i as it can be seen from the following formula :

$$pl_i(A) = \sum_{\substack{X \subseteq \Omega \\ X \cap A \neq \emptyset}} m_i(X), \forall A \in 2^\Omega. \quad (1)$$

Another useful concept is the **negation** (or complement) \bar{m}_i of a mass function m_i . The function \bar{m}_i is such that $\forall A \subseteq \Omega, \bar{m}_i(A) = m_i(\bar{A})$ with $\bar{A} = \Omega \setminus A$.

Besides, a mass function can be viewed as special kind of interval-valued probability distribution [9]. As part of the Transferable Belief Model, Smets [29] introduced the **pignistic transform** which is one way to retrieve a probability distribution with values inside the intervals specified by a mass function. This distribution is called the pignistic probability distribution $betp_i$ and it is obtained from m_i as follows : $\forall \omega \in \Omega$,

$$betp_i(\omega) = \sum_{A \subseteq \Omega, \omega \in A} \frac{1}{|A|} \frac{m_i(A)}{1 - m_i(\emptyset)}. \quad (2)$$

2.2 Mass function combination using α -junctions

In this subsection, a brief presentation of α -junctions [27] is proposed. Suppose two sources of information S_1 and S_2 have gathered pieces of evidence allowing them to define two mass functions m_1 and m_2 respectively. Evidential combination rules address the problem of aggregating these two functions into a single one synthesizing both of the initial evidence bodies. Let f be a combination operator for mass functions, *i.e.* $m_{12} = f(m_1, m_2)$ with m_{12} a mass function depending only on m_1 and m_2 . Such an operator is an **α -junction** if it possesses the following properties:

- Linearity¹: $\forall \lambda \in [0, 1], f(m, \lambda m_1 + (1 - \lambda) m_2) = \lambda f(m, m_1) + (1 - \lambda) f(m, m_2)$,
- Commutativity: $f(m_1, m_2) = f(m_2, m_1)$,
- Associativity: for any additional mass function $m_3, f(f(m_1, m_2), m_3) = f(m_1, f(m_2, m_3))$,
- Neutral element: $\exists m_e \mid \forall m, f(m, m_e) = m$,

¹The operator is linear on the vector space spanned by categorical mass functions but the output of the operator remains a mass function only in case of convex combination.

- Anonymity: for any mapping $\sigma : 2^\Omega \rightarrow 2^\Omega$ such that its restriction on Ω is a permutation and $\sigma(X) = \bigcup_{x \in X} \sigma(x)$ when $|X| > 1$, we have $f(m_1 \circ \sigma, m_2 \circ \sigma) = m_{12} \circ \sigma$,
- Context preservation: $pl_1(X) = 0$ and $pl_2(X) = 0 \implies pl_{12}(X) = 0$.

In short, α -junctions are thus linear combination rules that do not depend on the order in which pieces of evidence are processed. The justification behind these properties are detailed in [27]. In the same article, Smets proves that the neutral element m_e can only be either m_\emptyset or m_Ω . Depending on this, two sub-families arise: the α -disjunctive rules denoted by \odot^α and the α -conjunctive rules denoted by \ominus^α . For both of these families, Pichon and Dencœux [24] provided the following computation formulae: $\forall X \subseteq \Omega, \forall \alpha \in [0, 1]$

$$m_{1 \cap^\alpha 2}(X) = \sum_{(A \cap B) \cup (\bar{A} \cap \bar{B} \cap C) = X} m_1(A) m_2(B) \alpha^{|\bar{C}|} \bar{\alpha}^{|C|}, \quad (3)$$

$$m_{1 \cup^\alpha 2}(X) = \sum_{(A \Delta B) \cup (A \cap B \cap C) = X} m_1(A) m_2(B) \alpha^{|C|} \bar{\alpha}^{|\bar{C}|}, \quad (4)$$

with $\bar{\alpha} = 1 - \alpha$ and Δ the set symmetric difference. It is noteworthy that, if $\alpha = 1$, the classical conjunctive and disjunctive rules are retrieved. We denote these rules by $\odot = \odot^1$ and $\ominus = \ominus^1$. In addition, using \odot with a given function m_1 and a given categorical function m_B is tantamount to conditioning m_1 on B . The following notations will be used when combining m_B with a function m_1 :

- $m_{1|B} := m_1 \odot m_B$, this mass function is sometimes referred to as m_1 given B ,
- $m_{1 \cap^\alpha B} := m_1 \ominus^\alpha m_B$,
- $m_{1 \cup^\alpha B} := m_1 \odot^\alpha m_B$,
- $m_{1|^\alpha B}$ stands for the combination of m_1 and m_B using an α -junction when distinguishing conjunctive and disjunctive cases is unnecessary.

2.3 Meta-information and α -junctions

The interpretation of α -junctions is related to information items concerning the **truthfulness** of the sources S_1 and S_2 . In an information fusion context, such items are known as **meta-information** and truthfulness is a special kind of meta-information. Actually, several forms of truthfulness can be observed in practice but regarding α -junctions the following definition is retained: *S_i is untruthful if it supports the opposite of what it knows to be true.*

Depending on the truthfulness of sources, very different decisions can be made in the end, which accounts for the importance of taking meta-information into account in information fusion problems. In general, our knowledge about the

truthfulness of each source is imprecise and uncertain and it is therefore expressed as a mass function on a meta-domain \mathcal{T}_i . In [23], Pichon explains that an element $t_C^i \in \mathcal{T}_i$ is understood as the fact that S_i is truthful when it supports $\{\theta \in C\}$ and it is untruthful when it supports $\{\theta \in \overline{C}\}$. Let us provide a simple example when everything is deterministic:

Example 1. Suppose that $|\Omega| = 4$ and the meta-data concerning S_i is that t_C^i has probability 1 with $C = \{\omega_1, \omega_2\}$. If the source of information S_i delivers only one certain piece of evidence $\theta \in A = \{\omega_2, \omega_3\}$, then four different situations are encountered:

- The source gives support to ω_2 and can be trusted about ω_2 . We conclude that ω_2 is a possible value for θ .
- The source gives no support to ω_1 and can be trusted about ω_1 . We conclude that ω_1 is a not possible value for θ .
- The source gives support to ω_3 but cannot be trusted about ω_3 . We conclude that ω_3 is not a possible value for θ .
- The source gives no support to ω_4 but cannot be trusted about ω_4 . We conclude that ω_4 is a possible value for θ .

All in all, the testimony of the source is $\theta \in A$ but given the meta-data, the actual testimony is $\theta \in \overline{A\Delta C}$.

When considering a pair of sources $(S_1; S_2)$, meta-events belong to $\mathcal{T}_1 \times \mathcal{T}_2$. Pichon also proves that:

- for α -conjunctions, the underlying meta-information is that each meta-event *{either both sources are fully truthful or they both lie about \overline{C} }* = $\{(t_\Omega^1; t_\Omega^2); (t_C^1; t_C^2)\}$ has probability $\alpha^{|\overline{C}|} \overline{\alpha}^{|C|}$.
- for α -disjunctions, the underlying meta-information is that each meta-event *{one source is fully truthful while the other one lies at least about C }* = $\bigcup_{X \subseteq \overline{C}} \{(t_\Omega^1; t_X^2); (t_X^1; t_\Omega^2)\}$ has probability $\alpha^{|\overline{C}|} \overline{\alpha}^{|C|}$.

In particular, when $\alpha = 1$, the above probabilities are null whenever $C \neq \Omega$ in the conjunctive case and whenever $C \neq \emptyset$ in the disjunctive case. The meta-information thus reduces to:

- for the conjunctive rule, the event *{both sources are fully truthful}* has probability 1.
- for the disjunctive rule, the event *{at least one of the sources is fully truthful}* has probability 1.

Note that α -junctions are a particular case of a combination process introduced in [25] where a general framework for reasoning under various meta-information is formalized.

2.4 Belief functions and matrix calculus

Mass functions can be viewed as vectors belonging to the vector space \mathbb{R}^N with categorical mass functions as base vectors. Since mass functions sum to one, the set of mass functions is the simplex \mathcal{S} in that vector space whose vertices are the base vectors $\{m_A\}_{A \subseteq \Omega}$. This simplex is also called **mass space** [6]. In this paper, the following notations and conventions are used :

- Vectors are column vectors and are written in bold small letters. The operator **vec** maps a set function or a distribution to its vector form. For instance, $\mathbf{m}_i = \mathbf{vec}(m_i)$ is the mass vector corresponding to a mass function m_i . The length of mass vectors is N . The j_A th element of a mass vector \mathbf{m}_i is such that $\mathbf{m}_i(j_A) = m_i(A)$ with j_A the integer index of the set A according to the binary order. The binary order [28] is a common way to index elements of 2^Ω without supposing any order on Ω .
- Matrices are written in bold capital letters. They are square and their size is $N \times N$. A matrix can be represented by $\mathbf{X} = [X(i_A, j_B)]$, or alternatively by the notation $\mathbf{X} = [X(A, B)]$, $\forall A, B \in \Omega$. The row and column indexes i_A and j_B are those corresponding to the subsets A and B using the binary order.
- **I** is the identity matrix.
- **1** is the all-ones matrix.
- **J** is the matrix with null components except those on the anti-diagonal which are equal to 1. **J** is a permutation matrix reversing lines in case of right-handed product and reversing columns in case of left-handed product.

Matrix calculus as part of the theory of belief functions is especially interesting when it comes to mass function α -junctive combination. In [28] and [27], Smets shows that equation (3) and (4) can be written as a product between a matrix and a mass function vector. Let $\mathbf{K}_{1,\alpha}^\cap$ be a matrix such that $K_{1,\alpha}^\cap(A, B) = m_{1 \cap^\alpha B}(A)$ and $\mathbf{K}_{1,\alpha}^\cup$ a matrix such that $K_{1,\alpha}^\cup(A, B) = m_{1 \cup^\alpha B}(A)$. One has:

$$\mathbf{m}_{1 \cap^\alpha 2} = \mathbf{K}_{1,\alpha}^\cap \mathbf{m}_2, \quad (5)$$

$$\mathbf{m}_{1 \cup^\alpha 2} = \mathbf{K}_{1,\alpha}^\cup \mathbf{m}_2. \quad (6)$$

These matrices are also in one-to-one correspondence with the mass function m_1 . We call $\mathbf{K}_{1,\alpha}^\cap$ and $\mathbf{K}_{1,\alpha}^\cup$ the **α -specialization** and **α -generalization matrices** corresponding to m_1 . In general, all such matrices will be called **evidential matrices**. When it is not necessary to stress the dependency of evidential matrices on α and on the conjunctive/disjunctive cases, an evidential matrix is denoted by \mathbf{K}_1 for the sake of equation concision.

Each element of \mathbf{K}_1 represents the mass assigned to a set A after learning that $\{\theta \in B\}$: $K_1(A, B) = m_{1 \cap^\alpha B}(A)$. In other words, \mathbf{K}_1 does not only represent the current state of belief depicted by m_1 but also all reachable states from

m_1 through an α -junctive conditioning. From a geometric point of view [4], each column of an evidential matrix \mathbf{K}_1 corresponds to the vertex of a polytope \mathcal{P}_1 , called the **conditional subspace** of m_1 . Example 2 illustrates this latter remark.

Example 2. Let $|\Omega| = 2$ and $m_1 = \bar{\lambda}m_{\{\omega_1\}} + \lambda m_\Omega$ with $\lambda \in [0; 1]$ and $\bar{\lambda} = 1 - \lambda$. In the conjunctive case, we have :

$$\mathbf{K}_{1,\alpha}^\cap = \begin{pmatrix} \bar{\lambda}\alpha + \lambda & 0 & \bar{\lambda} & 0 \\ 0 & \bar{\lambda}\alpha + \lambda & 0 & \bar{\lambda} \\ \bar{\lambda}\bar{\alpha} & 0 & \lambda & 0 \\ 0 & \bar{\lambda}\bar{\alpha} & 0 & \lambda \end{pmatrix}.$$

The four column vectors of $\mathbf{K}_{1,\alpha}^\cap$ are (from left to right) : $\mathbf{m}_{1 \cap^\alpha \emptyset}$, $\mathbf{m}_{1 \cap^\alpha \{\omega_1\}}$, $\mathbf{m}_{1 \cap^\alpha \{\omega_2\}}$, and \mathbf{m}_1 . By definition the polytope \mathcal{P}_1 is the following subset of \mathcal{S} :

$$\left\{ m \in \mathcal{S} \mid m = \sum_{A \subseteq \Omega} \lambda_A \mathbf{m}_{1 \cap^\alpha A}, \sum_{A \subseteq \Omega} \lambda_A = 1, \lambda_A \in [0; 1] \right\}.$$

Any mass function $m \in \mathcal{P}_1$ is the result of the combination of m_1 with another mass function using a given α -junction. Evidential matrices are consequently relevant candidates for assessing dissimilarities between bodies of evidence in compliance with α -junctions.

Most importantly, if \mathbf{K}_1 and \mathbf{K}_2 are two evidential matrices and if \mathbf{K}_{12} denotes the matrix corresponding to the α -junction of m_1 with m_2 , then one has:

$$\mathbf{K}_{12} = \mathbf{K}_1 \mathbf{K}_2. \quad (7)$$

Moreover, the transpose of any evidential matrix \mathbf{K} is a stochastic matrix, meaning that all lines sum to one: ${}^t\mathbf{K} \mathbf{1} = \mathbf{1}$, with ${}^t\mathbf{K}$ the transpose matrix of \mathbf{K} . Finally, \mathbf{K}_A will denote the evidential matrix corresponding to the categorical mass function m_A .

3 Distances in evidence theory

In this section, we will first recall the definitions of some existing distances in the theory of belief functions. We will focus on various kinds of evidential dissimilarities without aiming at providing an exhaustive review. Afterward, new matrix-based evidential distances will be introduced in connection with α -junctions.

3.1 Vector-based evidential distances

A distance, or metric, provides a positive real value assessing the discrepancies between two elements. Let us first give a general definition of such an application when the compared vectors are mass functions:

Definition 1. Given a domain Ω and its related mass space \mathcal{S} , a mapping $d : \mathcal{S} \times \mathcal{S} \rightarrow [0, a]$ with $a \in \mathbb{R}^+$ is a **distance** between two mass functions m_1 and m_2 defined on Ω if the following properties hold:

- Symmetry : $d(m_1, m_2) = d(m_2, m_1)$,
- Definiteness : $d(m_1, m_2) = 0 \Leftrightarrow m_1 = m_2$,
- Triangle inequality : $d(m_1, m_2) \leq d(m_1, m_3) + d(m_3, m_2)$.

If the mapping fails to possess some of the above properties, then it degrades into unnormalized distance, dissimilarity or pseudo-distance. Only full metrics are able to provide a positive finite value that matches the intuitive notion of *gap*² between elements of a given space.

If $a \neq +\infty$, then the distance is bounded and if in addition $a = 1$, the distance is **normalized**. Provided that an evidential distance d is bounded, this distance can be normalized by dividing it with $\rho = \max_{m_1, m_2 \in \mathcal{S}} d(m_1, m_2)$. By definition, ρ is the **diameter** of \mathcal{S} . A result in convex geometry states that the diameter of a simplex is actually the length of its largest edge. Consequently, the following lemma holds:

Lemma 1. *Let d be a bounded evidential distance. Let \tilde{d} be the evidential distance defined by $\tilde{d}(m_1, m_2) = \frac{d(m_1, m_2)}{\rho}$ with $\rho = \max_{A, B \in 2^\Omega} d(m_A, m_B)$. Then, \tilde{d} is a normalized evidential distance.*

In evidence theory, two main families of mass function dissimilarities have been introduced, namely: direct and indirect dissimilarities. Direct measures are those defined on the product space $\mathcal{S} \times \mathcal{S}$ and thus complying with the geometrical interpretation of belief functions [6]. Concerning direct dissimilarities, the most widely used is Jousselme distance [16]. It is based on an inner product relying on a similarity matrix. This distance is given by:

$$d_J(m_1, m_2) = \sqrt{\frac{1}{2} t(\mathbf{m}_1 - \mathbf{m}_2) \mathbf{D} (\mathbf{m}_1 - \mathbf{m}_2)}, \quad (8)$$

with \mathbf{D} the Jaccard similarity matrix between focal elements. Its components are:

$$D(A, B) = \begin{cases} 1 & \text{if } A = B = \emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \text{otherwise} \end{cases}. \quad (9)$$

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

Many other evidential distances are defined similarly by substituting matrix \mathbf{D}

²This term was used by Frechet [13] in his early works on metric spaces, *i.e.* spaces endowed with a distance.

with another matrix evaluating the similarity between base vectors in different ways [10, 6]. Experimental material in [16] shows that these distances are highly correlated to d_J .

As for indirect dissimilarities, each mass function is first transformed into a vector belonging to a new space of representation $\mathcal{X} \subset \mathbb{R}^M$ related to another uncertainty theory. Afterwards, a distance measure is computed in this new representation space. When these transforms are lossy ($M < N$), indirect dissimilarities are not definite, hence none of them is a full metric. The most frequently used transform for indirect dissimilarities is the pignistic transform as proposed by Tessem [32] or Liu [20]. Tessem dissimilarity is defined as follows:

$$d_T(m_1, m_2) = \|\mathbf{betp}_1 - \mathbf{betp}_2\|_\infty, \quad (10)$$

with $\mathbf{betp}_i = \mathbf{vec}(betp_i)$ the vector corresponding to the pignistic probability distribution obtained from m_i .

In the same vein, Zouhal and Denceux [34] also define a dissimilarity measure based on the pignistic transform but they use the euclidean L^2 norm instead of the L^∞ norm:

$$d_{ZD}(m_1, m_2) = \|\mathbf{betp}_1 - \mathbf{betp}_2\|_2. \quad (11)$$

Other indirect dissimilarities [14] can be obtained in a space spanned by fuzzy set distributions. The vector-based distances cited in this subsection constitute a small sample of those found in the literature. However, this is a representative sample with respect to the diversity of evidential dissimilarities. They are therefore relevant candidates for a comparison with the new distances that will be introduced in the next subsection.

3.2 Matrix-based evidential distances

Since evidential matrices and mass functions are in one-to-one correspondence, assessing the distance between bodies of evidence can also be performed using distances between evidential matrices. The set of $N \times N$ matrices is denoted by \mathcal{M}_N and has the algebraic properties of a vector space as well. Consequently, matrix distances are not much different from vector distances. In this paper, we focus on matrix distances induced by matrix norms. A matrix norm is defined as follows:

Definition 2. A **matrix norm** $\|\cdot\|$ is a mapping defined on $\mathcal{M}_n \rightarrow \mathbb{R}^+$ satisfying the following conditions: \forall, \mathbf{A} and $\mathbf{B} \in \mathcal{M}_N$

1. $\|\mathbf{A}\| = 0 \Leftrightarrow \mathbf{A} = \mathbf{0}$,
2. $\|\lambda\mathbf{A}\| = |\lambda| \cdot \|\mathbf{A}\|$, for all $\lambda \in \mathbb{R}^+$,
3. $\|\mathbf{A} + \mathbf{B}\| \leq \|\mathbf{A}\| + \|\mathbf{B}\|$,

A matrix norm is sub-multiplicative, if in addition one has:

$$\|\mathbf{A} \mathbf{B}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|.$$

Any norm induces a distance defined as the norm of the difference between a pair of elements. Alleging that a matrix norm is bounded for evidential matrices, we thus introduce the following families of normalized evidential distances:

Definition 3. An α -specialization distance d is a mapping such that there exists a bounded matrix norm $\|\cdot\|$ and an α -conjunction \odot^α with:

$$\begin{aligned} d : \mathcal{S} \times \mathcal{S} &\rightarrow [0, 1], \\ m_1 \times m_2 &\rightarrow \frac{1}{\rho} \|\mathbf{K}_{1,\alpha}^\cap - \mathbf{K}_{2,\alpha}^\cap\|. \end{aligned} \quad (12)$$

$\mathbf{K}_{i,\alpha}^\cap$ is the α -specialization matrix corresponding to m_i and $\rho = \max_{A,B \in 2^\Omega} \|\mathbf{K}_{A,\alpha}^\cap - \mathbf{K}_{B,\alpha}^\cap\|$ is a normalization factor.

Definition 4. An α -generalization distance d is a mapping such that there exists a bounded matrix norm $\|\cdot\|$ and an α -disjunction \odot^α with:

$$\begin{aligned} d : \mathcal{S} \times \mathcal{S} &\rightarrow [0, 1], \\ m_1 \times m_2 &\rightarrow \frac{1}{\rho} \|\mathbf{K}_{1,\alpha}^\cup - \mathbf{K}_{2,\alpha}^\cup\|. \end{aligned} \quad (13)$$

$\mathbf{K}_{i,\alpha}^\cup$ is the α -generalization matrix corresponding to m_i and $\rho = \max_{A,B \in 2^\Omega} \|\mathbf{K}_{A,\alpha}^\cup - \mathbf{K}_{B,\alpha}^\cup\|$ is a normalization factor.

The family of α -specialization distances is an extension of the family introduced in [21] which corresponds to the $\alpha = 1$ case. Among existing matrix norms, we focus in this article on the most frequently used ones, namely: L^k norms and operator norms.

L^k matrix norms are also known as entry-wise norms. Since matrices are elements of the vector space \mathcal{M}_N , the definition of L^k matrix norms is the following:

$$\|\mathbf{A}\|_k = \left(\sum_{1 \leq j \leq n} \sum_{1 \leq i \leq n} |\mathbf{A}_{ij}|^k \right)^{\frac{1}{k}}. \quad (14)$$

Both L^k vector norms and L^k matrix norms are denoted by $\|\cdot\|_k$. They are easily distinguished since vectors are in small letters and matrices are in capital letters.

The k -operator norm $\|\cdot\|_{opk}$, also known as induced norm, is defined for any matrix $\mathbf{A} \in \mathcal{M}_N$ as follows:

$$\|\mathbf{A}\|_{opk} = \max_{\mathbf{x} \in \mathbb{R}^N, \mathbf{x} \neq \mathbf{0}} \frac{\|\mathbf{Ax}\|_k}{\|\mathbf{x}\|_k} = \max_{\mathbf{x} \in \mathbb{R}^N, \|\mathbf{x}\|_k=1} \|\mathbf{Ax}\|_k. \quad (15)$$

with $\|\cdot\|_k$ the classical L^k vector norm. In particular, the 1-operator norm writes as:

$$\|\mathbf{A}\|_{op1} = \max_{1 \leq j \leq n} \sum_{1 \leq i \leq n} |\mathbf{A}_{ij}|. \quad (16)$$

It can be noted that any evidential matrix \mathbf{K} is such that $\|\mathbf{K}\|_{op1} = 1$. In the sequel, d_k denotes the α -specialization or α -generalization distance relying on the L^k matrix norm. The notation d_{opk} is used when the distance relies on the k -operator norm. In the same fashion as for evidential matrices, the value of α and the conjunctive or disjunctive nature of these distances are given in exponent when necessary.

A first result concerning the newly introduced distances is that there is a duality between α -specialization distances and α -generalization distances that has its roots in the De Morgan relations proved in [28]. The following proposition formalizes this duality.

Proposition 1. *Suppose $\alpha \in [0, 1]$. Let d^\cap be an α -specialization distance with respect to the α -conjunctive rule \odot^α and relying on an operator or L^k matrix norm. Let d^\cup be an α -generalization distance with respect to the α -disjunctive rule \ominus^α and relying on the same norm. For any mass functions m_1 and m_2 on a domain Ω , one has:*

$$d^\cap(m_1, m_2) = d^\cup(\overline{m}_1, \overline{m}_2). \quad (17)$$

Proof. Let \mathbf{K}_i^\cap and \mathbf{K}_i^\cup be the α -specialization matrix and the α -generalization matrix of m_i with $i \in \{1; 2\}$. Let $\overline{\mathbf{K}}_i^\cap$ and $\overline{\mathbf{K}}_i^\cup$ be the α -specialization matrix and the α -generalization matrix of \overline{m}_i with $i \in \{1; 2\}$. According to theorem 12.1 in [28], we have $\mathbf{K}_i^\cap = \mathbf{J}\overline{\mathbf{K}}_i^\cup\mathbf{J}$ with \mathbf{J} the binary anti-diagonal matrix. It follows that:

$$\begin{aligned} d^\cap(m_1, m_2) &= \frac{1}{\rho} \|\mathbf{K}_1^\cap - \mathbf{K}_2^\cap\|, \\ &= \frac{1}{\rho} \|\mathbf{J}\overline{\mathbf{K}}_1^\cup\mathbf{J} - \mathbf{J}\overline{\mathbf{K}}_2^\cup\mathbf{J}\|, \\ &= \frac{1}{\rho} \|\mathbf{J}(\overline{\mathbf{K}}_1^\cup - \overline{\mathbf{K}}_2^\cup)\mathbf{J}\|. \end{aligned}$$

Since \mathbf{J} is a permutation matrix and that any L^k matrix norm or operator norm computation is invariant to column or line relabeling, one has $\|\mathbf{J}(\overline{\mathbf{K}}_1^\cup - \overline{\mathbf{K}}_2^\cup)\mathbf{J}\| = \|\overline{\mathbf{K}}_1^\cup - \overline{\mathbf{K}}_2^\cup\| = \rho d^\cup(\overline{m}_1, \overline{m}_2)$. \square

Proposition 1 sheds light on ties between α -specialization distances and α -generalization distances. When $\alpha \in \{0; 1\}$, it appears that these ties are stronger as illustrated by Lemma 2:

Lemma 2. *Let $d^{\cap, \alpha}$ be an α -specialization distance with respect to the α -conjunctive rule \odot^α and relying on an operator or L^k matrix norm. Let $d^{\cup, \alpha}$ be an α -generalization distance with respect to the α -disjunctive rule \ominus^α and relying on the same norm. For any mass functions m_1 and m_2 on a domain Ω , one has:*

$$d^{\cap, 0} = d^{\cup, 0}, \quad (18)$$

$$d^{\cap, 1} = d^{\cup, 1}. \quad (19)$$

The proof of Lemma 2 is given in appendix A. This result shows that, for extreme values of α , α -specialization distances and α -generalization distances are identical. Consequently, the underlying meta-information is treated in the same way in these special cases. The following example shows that this is not true when $\alpha \in]0, 1[$:

Example 3. Suppose Ω is a domain and X a subset of Ω . Let m be a mass function such that:

$$m = 0.3m_X + 0.5m_{\overline{X}} + 0.2m_\Omega.$$

Figure 1 shows α -specialization and α -generalization distances relying on the L^1 matrix norm between m and m_X when $|\Omega| = 3$ and $|X| = 2$.

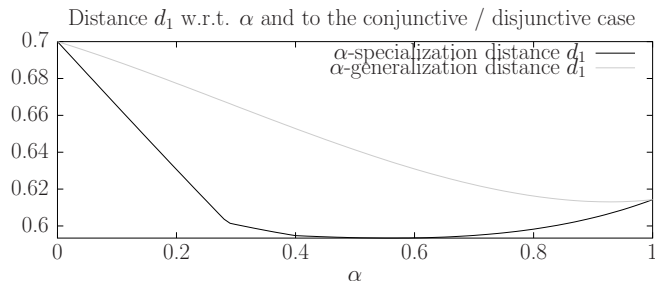


Figure 1: Different α -specialization and α -generalization distances relying on the L^1 matrix norm. These distances are computed between two given mass functions m_1 and m_2 such that $m_1 = m_X$ and $m_2 = 0.3m_X + 0.5m_{\overline{X}} + 0.2m_\Omega$, with $|\Omega| = 3$ and $|X| = 2$.

Besides, proposition 1 allows us to anticipate the fact that if an α -specialization distance satisfies a given property then so does its α -generalization counterpart. Concerning evidential distance properties, we argued in [21] that, on top of metric properties, two other properties are especially interesting. The definitions and justifications of each of them are given respectively in section 4 and 5. We will also examine what α -specialization distances and α -generalization distances may satisfy these properties.

4 Consistency with α -junctions

One way to formalize the consistency of an evidential distance with a combination rule is given by the following definition:

Definition 5. Let \odot be a combination rule and d an evidential distance. d is said to be **consistent** with respect to \odot if for any mass functions m_1, m_2 and m_3 on Ω :

$$d(m_1 \odot m_3, m_2 \odot m_3) \leq d(m_1, m_2). \quad (20)$$

Suppose each mass function m can be decomposed into elementary pieces of evidence such that their aggregation using a combination rule gives m . For instance, Shafer [26] introduced the canonical decomposition under Dempster's rule of a particular class of mass functions into simple mass function sets. Using a consistent evidential distance, mass functions are all the closer as their decompositions involve identical elementary pieces of evidence.

4.1 Results on the consistency of matrix-based evidential distances

We provide several results concerning the consistency of α -specialization distances and α -generalization distances. The first one deals with these distances when the 1-operator norm is used:

Proposition 2. *Any α -specialization or α -generalization distance d_{op1} defined using the 1-operator norm is consistent with its corresponding α -junctive combination rule.*

Proof. Suppose m_1, m_2 and m_3 are three mass functions defined on Ω . $\mathbf{K}_1, \mathbf{K}_2$ and \mathbf{K}_3 are their respective evidential matrices with respect to a given α -junction denoted by \odot^α . The 1-operator-norm has the sub-multiplicative property, *i.e.* for all matrices \mathbf{A} and \mathbf{B} , one has:

$$\|\mathbf{AB}\|_{op1} \leq \|\mathbf{A}\|_{op1} \cdot \|\mathbf{B}\|_{op1}$$

One can thus write:

$$\|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{K}_3\|_{op1} \leq \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} \cdot \|\mathbf{K}_3\|_{op1},$$

Given that $\|\mathbf{K}_3\|_{op1} = 1$, we have:

$$\begin{aligned} \|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{K}_3\|_{op1} &\leq \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1}, \\ \Leftrightarrow d_{op1}(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3) &\leq d_{op1}(m_1, m_2). \end{aligned}$$

By definition, this latter inequality means that distance d_{op1} is consistent with rule \odot^α . \square

Another result holds when α -specialization or α -generalization distances are defined using the L^1 matrix norm. To prove this result, it is first necessary to introduce the following lemma:

Lemma 3. *Suppose m_1 and m_2 are two mass functions defined on Ω and A and B are two subsets such that $A \subseteq B \subseteq \Omega$. Then, the following properties hold for any α -conjunctive rule \odot^α and any α -disjunctive rule \oplus^α*

$$\|\mathbf{m}_{1 \cap^\alpha A} - \mathbf{m}_{2 \cap^\alpha A}\|_1 \leq \|\mathbf{m}_{1 \cap^\alpha B} - \mathbf{m}_{2 \cap^\alpha B}\|_1, \quad (21)$$

$$\|\mathbf{m}_{1 \cup^\alpha A} - \mathbf{m}_{2 \cup^\alpha A}\|_1 \geq \|\mathbf{m}_{1 \cup^\alpha B} - \mathbf{m}_{2 \cup^\alpha B}\|_1. \quad (22)$$

The proof of Lemma 3 is given in appendix B. From this lemma, we deduce the following corollary:

Corollary 1. *Suppose m_1 and m_2 are two mass functions defined on Ω . Let d_{op1} be an α -specialization or α -generalization distance defined using the 1-operator norm. We have:*

$$d_{op1}(m_1, m_2) = \frac{1}{\rho} \|\mathbf{m}_1 - \mathbf{m}_2\|_1. \quad (23)$$

This means that the 1-operator distance d_{op1} is the L^1 distance between mass vectors for any α . The proof of corollary 1 is given in appendix C. This corollary also implies that $\rho = 2$ for d_{op1} . Equipped with these preliminary results, we are now able to provide the following proposition regarding the consistency of distances defined using the L^1 norm:

Proposition 3. *Any α -specialization or α -generalization distance d_1 defined using the L^1 matrix norm is consistent with its corresponding α -junctive combination rule.*

Proof. Suppose m_1, m_2 and m_3 are three mass functions defined on Ω . Suppose $\mathbf{K}_1, \mathbf{K}_2$ and \mathbf{K}_3 are their respective evidential matrices with respect to an α -junction denoted by \odot^α . The L^1 norm of a matrix is the sum over all columns of the L^1 norms of its column vectors:

$$\|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{K}_3\|_1 = \sum_{A \subseteq \Omega} \|\mathbf{m}_{13|\alpha A} - \mathbf{m}_{23|\alpha A}\|_1, \quad (24)$$

with $m_{i3|\alpha A} = m_i \odot^\alpha m_A \odot^\alpha m_3$. Besides, according to proposition 2, d_{op1} is consistent with \odot^α . Inequality (20) thus applies, and after multiplying both sides of this inequality by ρ , we have that for any $A \subseteq \Omega$:

$$\|(\mathbf{K}_1|\alpha A - \mathbf{K}_2|\alpha A) \mathbf{K}_3\|_{op1} \leq \|\mathbf{K}_1|\alpha A - \mathbf{K}_2|\alpha A\|_{op1},$$

with $\mathbf{K}_i|\alpha A$ the evidential matrix corresponding to $m_i|\alpha A$ with respect to \odot^α . Now applying corollary 1 on each side of the above inequality gives:

$$\|\mathbf{m}_{13|\alpha A} - \mathbf{m}_{23|\alpha A}\|_1 \leq \|\mathbf{m}_1|\alpha A - \mathbf{m}_2|\alpha A\|_1. \quad (25)$$

Let us now use this inequality in equation (24):

$$\begin{aligned} \|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{K}_3\|_1 &\leq \sum_{A \subseteq \Omega} \|\mathbf{m}_1|\alpha A - \mathbf{m}_2|\alpha A\|_1, \\ \|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{K}_3\|_1 &\leq \|(\mathbf{K}_1 - \mathbf{K}_2)\|_1. \end{aligned}$$

Finally, dividing both sides of the above inequality by ρ gives:

$$d_1(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3) \leq d_1(m_1, m_2).$$

By definition, this latter inequality means that distance d_1 is consistent with rule \odot^α . \square

One last result is available for α -specialization or α -generalization distances defined using the L^∞ matrix norm:

Proposition 4. *Any α -specialization or α -generalization distance d_∞ defined using the L^∞ matrix norm is consistent with its corresponding α -junctive combination rule.*

Proof. Suppose m_1 , m_2 and m_3 are three mass functions defined on Ω . \mathbf{K}_1 , \mathbf{K}_2 and \mathbf{K}_3 are their respective evidential matrices with respect to an α -junction \odot^α . The L^∞ norm of a matrix is the max of the L^∞ norms of its column vectors. Since a column vector of \mathbf{K}_i writes as $\mathbf{m}_{i|^\alpha B}$ with $B \subseteq \Omega$, there exists a subset X such that :

$$\begin{aligned} \|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{K}_3\|_\infty &= \|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{m}_{3|^\alpha X}\|_\infty, \\ &= \left\| (\mathbf{K}_1 - \mathbf{K}_2) \sum_{Y \subseteq \Omega} m_{3|^\alpha X}(Y) \mathbf{m}_Y \right\|_\infty, \\ &\leq \sum_{Y \subseteq \Omega} m_{3|^\alpha X}(Y) \|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{m}_Y\|_\infty. \end{aligned}$$

Again, the L^∞ norm of $\mathbf{K}_1 - \mathbf{K}_2$ is the max of the L^∞ norms of its columns vectors : $\max_{Y \subseteq \Omega} \|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{m}_Y\|_\infty = \|(\mathbf{K}_1 - \mathbf{K}_2)\|_\infty$. Each term in the previous inequation is maximized by $\|(\mathbf{K}_1 - \mathbf{K}_2)\|_\infty$ which gives:

$$\begin{aligned} \|(\mathbf{K}_1 - \mathbf{K}_2) \mathbf{K}_3\|_\infty &\leq \sum_{Y \subseteq \Omega} m_{3|^\alpha X}(Y) \|(\mathbf{K}_1 - \mathbf{K}_2)\|_\infty, \\ &\leq \|(\mathbf{K}_1 - \mathbf{K}_2)\|_\infty. \end{aligned}$$

After normalization, the above inequation gives $d_\infty(m_1 \odot^\alpha m_3, m_2 \odot^\alpha m_3) \leq d_\infty(m_1, m_2)$. \square

The following subsection investigates the inconsistency of distances d_J , d_2 , d_{op2} and $d_{op\infty}$ with respect to α -junctions through a numerical experiment.

4.2 Experiments on consistency

This section contains experiments illustrating the (in)consistency of several evidential distances with respect to α -conjunctive and α -disjunctive rules. For each distance and each value of α , one iteration of this experiment consists in picking three random mass functions and check if inequality (20) is verified. The number of times that the property is verified over the number of iterations gives the **consistency rate** of the distance for a given α .

In order to provide such rates, it is necessary to generate mass functions randomly. It is sufficient to draw simple mass functions because cases of inconsistency are more frequent with such functions. Random simple mass functions are drawn uniformly using an algorithm presented in [2] and applied to simple mass function sub-simplices.

In figure 2, consistency rates for α -conjunctive rules and several evidential distances are shown. For this experiment, $1e4$ iterations were used. Figure 3 shows the same results for α -disjunctive rules.

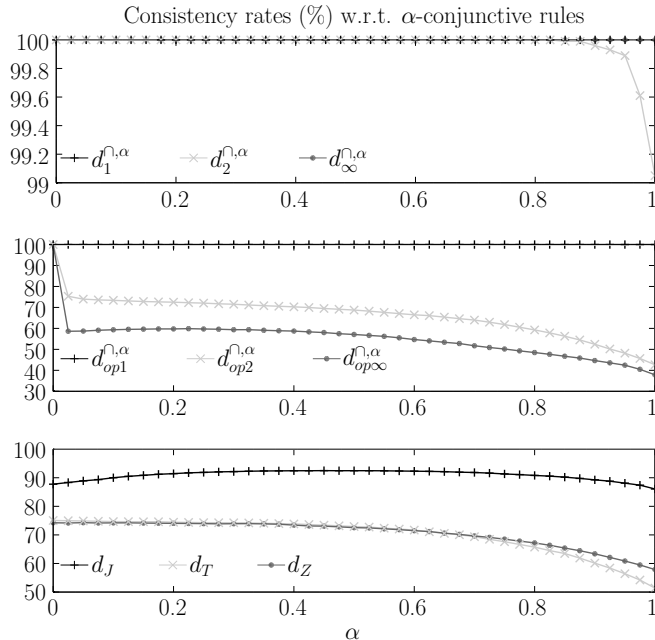


Figure 2: Consistency rates of several evidential distances with α -conjunctive rules with respect to parameter α .

As expected, the rates of d_{op1} , d_1 and d_∞ are 100% in both the conjunctive and disjunctive cases. A rate under 100% is sufficient to prove the inconsistency of a distance. We can therefore conclude that d_2 , d_{op2} and $d_{op\infty}$ are inconsistent when $\alpha \neq 0$. It can be conjectured that they are consistent when $\alpha = 0$. The experiment also proves the inconsistency of d_J , d_T and d_Z with any α -junction except for the disjunctive case when $\alpha = 0$. Their consistency in this latter case may also be conjectured.

5 Poset structure and evidential matrix-based metrics

In [17], Jousselme and Maupin shed light on the fact that evidential distances should not process mass functions as if they were any vectors living in the mass space \mathcal{S} . Indeed, the power set 2^Ω endowed with the inclusion partial order \subseteq is a poset, therefore, $A \subseteq B$ should induce a similar relation between the base vectors \mathbf{m}_A and \mathbf{m}_B with consequences on distance values.

In [21], we formalized this intuitive principle into the *structure property* characterizing the existence of an order-isomorphism between an evidential distance

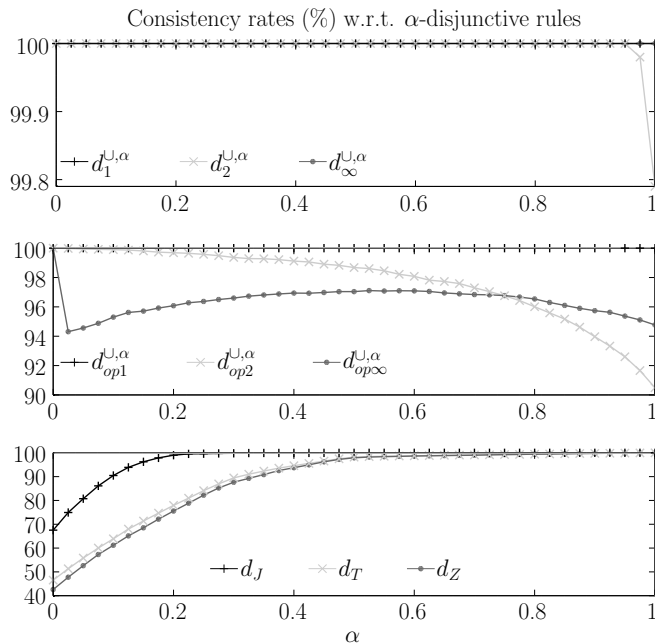


Figure 3: Consistency rates of several evidential distances with α -disjunctive rules with respect to parameter α .

between categorical mass function and a relevant set distance³. Experiments in [21] show that this property is sufficient but not necessary to guarantee that the poset structure is correctly taken into account. In this article, we thus introduce a new property called **\subseteq -compatibility property**:

Definition 6. An evidential distance d is said to be **\subseteq -compatible** if its restriction to categorical mass functions is not the trivial⁴ set distance d_{triv} and if $\forall A, B, C \subseteq \Omega$ such that $A \subseteq B \subseteq C$, one has:

$$d(m_A, m_B) \leq d(m_A, m_C). \quad (26)$$

This property ensures that evidential distances between categorical mass functions have similar dynamics as relevant set distances like the Jaccard distance or the Hamming set distance.

5.1 Results on the compatibility of matrix-based evidential distances with the inclusion partial order

Among α -specialization or α -generalization distances, let us first focus on d_{op1} and d_∞ :

³A set distance is a metric on $(2^\Omega, \subseteq)$.

⁴For any sets A and B , the trivial metric equals 1 whenever $A \neq B$.

Proposition 5. *Let Ω be a domain and $A, B \subseteq \Omega$. Any α -specialization or α -generalization distance d_{op1} or d_∞ defined respectively using the 1-operator or the L^∞ matrix norm is \subseteq -incompatible. Moreover, for any categorical mass functions m_A and m_B , we have:*

$$d_{op1}(m_A, m_B) = d_\infty(m_A, m_B) = d_{triv}(A, B). \quad (27)$$

Proof. Since d_{op1} and d_∞ are metrics, we have $d_{op1}(m_A, m_B) = d_\infty(m_A, m_B) = 0$ when $A = B$.

Let us now focus on the $A \neq B$ case. From corollary 1, we know that $d_{op1}(m_A, m_B) = \frac{1}{\rho} \|\mathbf{m}_A - \mathbf{m}_B\|_1$. Distances d_{op1} and d_∞ share the same normalization constant: $\rho = 2$. If $A \neq B$, then $\|\mathbf{m}_A - \mathbf{m}_B\|_1 = 2$ and thus $d_{op1}(m_A, m_B) = 1$. In addition, for categorical mass functions, we have $\|\mathbf{m}_A - \mathbf{m}_B\|_1 = \|\mathbf{m}_A - \mathbf{m}_B\|_\infty$, hence:

$$\begin{aligned} d_{op1}(m_A, m_B) &= \frac{1}{2} \|\mathbf{m}_A - \mathbf{m}_B\|_\infty, \\ \Rightarrow d_{op1}(m_A, m_B) &\leq \frac{1}{2} \|\mathbf{K}_A - \mathbf{K}_B\|_\infty, \\ \Leftrightarrow d_{op1}(m_A, m_B) &\leq d_\infty(m_A, m_B). \end{aligned} \quad (28)$$

Since d_∞ is normalized, we deduce that $d_\infty(m_A, m_B) = 1$. \square

The above result shows that the definition of the \subseteq -compatibility property makes sense because d_{op1} is actually equivalent to a vector distance which does not grasp the poset structure of 2^Ω at all.

The following lemma shows that the restriction of d_1 to the set of categorical mass functions is not the trivial distance:

Lemma 4. *Suppose m_A and m_B are two categorical mass functions on A and B , two subsets of Ω . Let d_1 be an L^1 based distance with respect to a given α -junction. $\forall \alpha \in [0, 1]$, the following result for d_1 holds :*

$$d_1(m_A, m_B) = \frac{2N}{\rho} \left(1 - \frac{\alpha^{\max\{|A \setminus B|; |B \setminus A|\}}}{2^{|A \Delta B|}} \right). \quad (29)$$

The proof of this lemma is found in appendix D. This proof also shows that $\rho = 2(N - \alpha^n)$ for d_1 . Equipped with this lemma, one is now able to prove the following proposition:

Proposition 6. *Any α -specialization or α -generalization distance d_1 defined using the L^1 matrix norm is \subseteq -compatible provided that $\alpha \neq 0$.*

Proof. Under the $\alpha = 0$ case, we have $d_1(m_A; m_B) = d_{triv}(A, B)$ and d_1 is \subseteq -incompatible.

Let us now suppose that $\alpha > 0$. For any sets $A, B \subseteq \Omega$ such that $A \subseteq B$, we have $\max\{|A \setminus B|; |B \setminus A|\} = |B \setminus A|$ and $|A \Delta B| = |B \setminus A|$, which implies $d_1(m_A, m_B) = \frac{2N}{\rho} \left(1 - \left(\frac{\alpha}{2}\right)^{|B \setminus A|} \right)$. Now if C is such that $A \subseteq B \subseteq C$, then $|B \setminus A| \leq |C \setminus A|$ which is equivalent to $d_1(m_A, m_B) \leq d_1(m_A, m_C)$. \square

The next subsection gives more details on the compatibility of other evidential distances with the inclusion partial order thanks to a numerical experiment.

5.2 Experiments on compatibility with the inclusion partial order

The experiments contained in this section are meant to determine what evidential distances are compliant with inequation (26) when categorical mass functions are compared. To that end, let us study the following toy-example:

Example 4. This example is inspired from an example presented in [16] and re-used in [14, 21]. Let m_X and m_{A_i} be two categorical mass functions defined on the domain $\Omega = \{\omega_i\}_{i=1}^n$, with $X = \{\omega_1, \omega_2, \omega_3\}$ and $n = |\Omega| = 7$. The subset A_i varies by successive inclusions of a singleton at each step starting from \emptyset and reaching Ω at the last computation step:

i : computation step	focal element A_i
0	\emptyset
1	$\{\omega_1\}$
2	$\{\omega_1, \omega_2\}$
3	$\{\omega_1, \omega_2, \omega_3\} = X$
4	$\{\omega_1, \dots, \omega_4\}$
...	...
7	Ω

The values of several evidential distances are shown in figure 4 for each step i . Note that it is unnecessary to include α -generalization distances in this experiment because we know from proposition 1 that they behave similarly as α -specialization distances when comparing categorical mass functions.

A \subseteq -compatible distance must be decreasing when $i \in \{0; \dots; 3\}$ and increasing when $i \in \{3; \dots; 7\}$. To this regard, d_2 (when $\alpha = 0.25$), d_{op2} (when $\alpha = 0.5$) and $d_{op\infty}$ (when $\alpha = 0.25$) are \subseteq -incompatible. It can be conjectured that any matrix-based evidential distance between categorical mass function is the trivial distance when $\alpha = 0$.

As for vector distances, d_Z is also \subseteq -incompatible while it can be proved that d_T and d_J are both \subseteq -compatible.

6 Meta-information influence on distances

This section highlights the behavior of α -specialization and α -generalization distances in the context of a meta-information dependent information fusion problem. We limit ourselves to meta-information items pertaining to α -junctions as described in subsection 2.3, *i.e.* source states.

When the source states are completely unknown, the behavior-based correction introduced by Pichon *et al.* [25] gives that mass functions delivered by the sources should be replaced with vacuous ones. Any evidential distance is thus null in such a case. Given that the sources are equally poorly informative, this result makes sense.

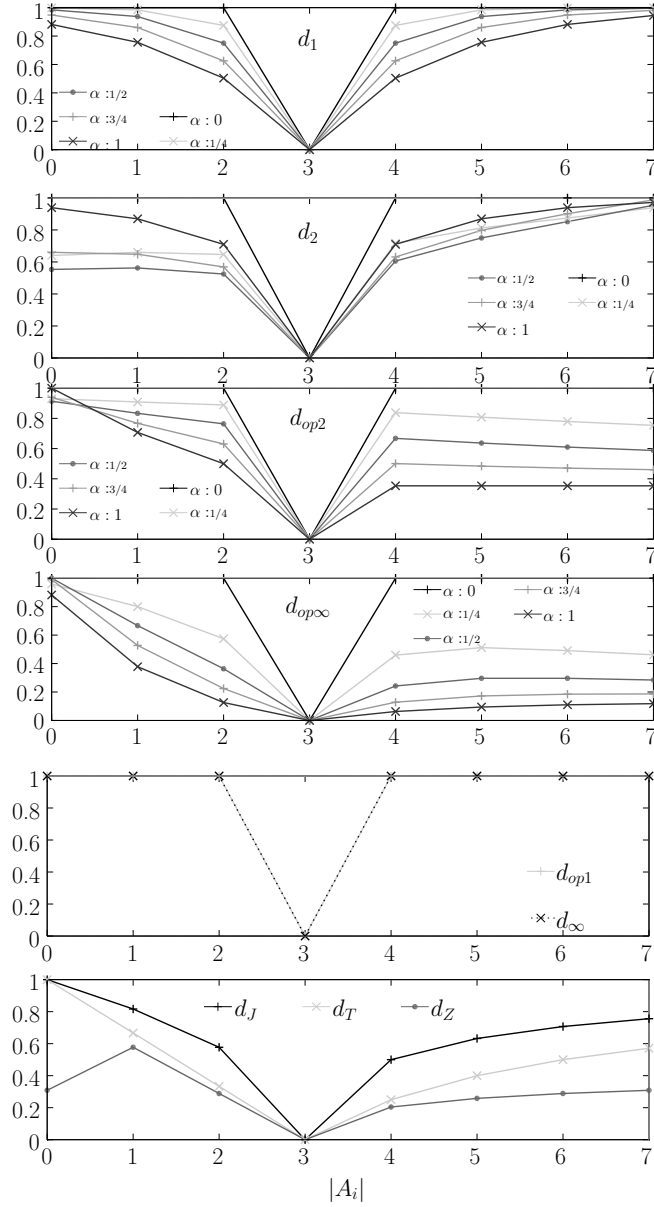


Figure 4: Evidential distances and dissimilarities between two categorical mass functions: m_X and m_{A_i} defined on Ω , with $|\Omega| = 7$ and $|X| = 3$. The set A_i varies by successive inclusions from \emptyset to Ω and $A_3 = X$. The matrix-based evidential distances involved in this experiment are α -specialization distances and five values of α are considered: $\{0; \frac{1}{4}; \frac{1}{2}; \frac{3}{4}; 1\}$.

Let us now examine the opposite situation where meta-information is certain and precise, *i.e.* the meta-events $\{\text{source } S_1 \text{ is in state } t_C^1\}$ and $\{\text{source } S_2 \text{ is in state } t_D^2\}$ are sure. In this case, the same correction mechanism can be used on each mass function to make them deliver only righteous pieces of evidence. Indeed, back to example 1, we came to the conclusion that a categorical mass function m_A delivered by a source in state t_C^i should be replaced with the function $m_{\overline{A\Delta C}}$. Now, more generally, if $m_{1|t_C^1}$ denotes the mass function that S_1 should have delivered, then we deduce:

$$m_{1|t_C^1}(Y) = \sum_{\substack{X \subseteq \Omega \\ Y = \overline{C\Delta X}}} m_1(X) = m_1(\overline{C\Delta Y}), \quad (30)$$

because $X = \overline{C\Delta Y} \Leftrightarrow Y = \overline{C\Delta X}$.

Using this mechanism, one can compute distances between corrected mass functions without minding meta-information. Although this mechanism is simple and easy to implement, it is however unlikely that such steady meta-information is available in practice.

A more realistic situation is when both sources lie in the same way but one does not know which lie is committed. For instance, many real-world systems possess redundant sensors which may be malfunctioning in the same way because they undergo the same experimental conditions. In this particular situation, a relevant mass function comparison is delivered by some α -specialization and α -generalization distances in spite of the meta-information imprecision. Indeed, we have the following result:

Proposition 7. *Let m_1 and m_2 be two mass functions on a domain Ω delivered by two sources S_1 and S_2 that are in states t_C^1 and t_C^2 . If d is a specialization distance relying on an L^k matrix norm or on the 1-operator norm, then we have:*

$$d(m_1, m_2) = d\left(m_{1|t_C^1}, m_{2|t_C^2}\right), \quad (31)$$

with $m_{1|t_C^1}$ and $m_{2|t_C^2}$ the mass functions that would have been delivered by S_1 and S_2 if they were both fully truthful.

Proof. Let Ω be a domain and $B, C \subseteq \Omega$. Let σ_B denote a mapping such that:

$$\begin{aligned} \sigma_B : 2^B &\rightarrow 2^B, \\ X &\rightarrow \overline{C\Delta X} \cap B. \end{aligned}$$

This mapping is idempotent: $\sigma_B^{-1}(Y) = \sigma_B(Y)$. Indeed, for any $X \subseteq B$, we

have:

$$\begin{aligned}
\sigma_B(\sigma_B(X)) &= \overline{(\overline{X\Delta\overline{C}} \cap B) \Delta C} \cap B, \\
&= ((\overline{X\Delta\overline{C}} \cap B \cap C) \cup \\
&\quad (\overline{X\Delta\overline{C}} \cap B \cap \overline{C})) \cap B, \\
&= (\overline{X\Delta\overline{C}} \cap B \cap C) \cup \\
&\quad ((X\Delta C) \cup \overline{B}) \cap \overline{C} \cap B, \\
&= (X \cap C \cap B \cap C) \cup \\
&\quad (\overline{X} \cap \overline{C} \cap B \cap C) \cup \\
&\quad (X\Delta C) \cap \overline{C} \cap B, \\
&= (X \cap C) \cup (X \cap \overline{C} \cup \overline{X} \cap C) \cap \overline{C} \cap B, \\
&= (X \cap C) \cup (X \cap \overline{C}), \\
&= X.
\end{aligned}$$

σ_B is thus a permutation on 2^B .

From equation (30), a corrected mass function $m_{i|t_C^i}$ is obviously equal to $m_i \circ \sigma_\Omega$. This is sufficient to conclude for the 1-operator norm because this norm is actually a vector norm whose computation is invariant with respect to component permutations. For L^k matrix norms, one must first remark that $\forall A, B, X \subseteq \Omega$,

$$X \cap B = A \cap B \Leftrightarrow \sigma_\Omega(X) \cap B = \sigma_\Omega(A) \cap B.$$

It follows that $m_{i|B}(A \cap B) = m_{i|t_C^i, B}(\sigma_B(A))$, with $m_{i|t_C^i, B} = m_{i|t_C} \odot m_B$. Let \mathbf{S}_i and $\mathbf{S}_{i|t_C^i}$ be the specialization matrices corresponding to m_i and $m_{i|t_C^i}$ respectively. Remembering that $\mathbf{m}_{i|B}$ and $\mathbf{m}_{i|t_C^i, B}$ are the j_B^{th} (in the binary order sense) column vectors of \mathbf{S}_i and $\mathbf{S}_{i|t_C^i}$, it appears that each column of $\mathbf{S}_{i|t_C^i}$ is obtained by permuting non-null elements of the same column in \mathbf{S}_i . Since L^k matrix norm computation is invariant to matrix component permutations, one has :

$$\begin{aligned}
\|\mathbf{S}_{1|t_C^1} - \mathbf{S}_{2|t_C^2}\|_k &= \|\mathbf{S}_1 - \mathbf{S}_2\|_k, \\
\Leftrightarrow d_k(m_1, m_2) &= d_k(m_{1|t_C^1}, m_{2|t_C^2}).
\end{aligned}$$

□

Proposition 7 is interesting in the sense that, when sources are corrupted in same way, the same distance value is obtained as if they were delivering truthful pieces of evidence. This property is illustrated in the following example:

Example 5. The same mass functions as in example 3 are used: $m_1 = m_X$ and $m_2 = 0.3m_X + 0.5m_{\overline{X}} + 0.2m_\Omega$. However, we chose here $|\Omega| = 7$ and $|X| = 3$. Suppose C_i is a subset varying by successive inclusions from \emptyset to Ω and $C_3 = X$. Figure 5 displays distances between $m_{1|t_{C_i}^1}$ and $m_{2|t_{C_i}^2}$.

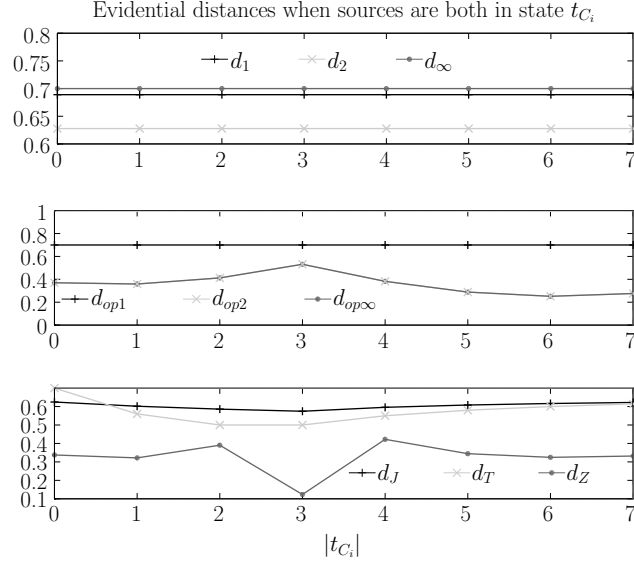


Figure 5: Different evidential distances between same-state-sources. These distances are computed between two given mass functions $m_1|_{t_{C_i}^1}$ and $m_2|_{t_{C_i}^2}$ such that $m_1 = m_X$ and $m_2 = 0.3m_X + 0.5m_{\overline{X}} + 0.2m_\Omega$, with $|\Omega| = 7$ and $|X| = 3$. The state set C_i varies by successive inclusions from \emptyset to Ω and $A_3 = X$.

Significant discrepancies are observed for vector distances as well as for d_{op2} and $d_{op\infty}$ showing that their distance values are irrelevant if meta-knowledge cannot be filtered.

7 Conclusion

In the theory of belief functions, evidence is summarized through a mass function. Since mass functions are in one-to-one correspondence with several matrices, two bodies of evidence can be compared by computing the norm of the difference between such matrices. By investigating several matrix norms, infinite families of evidential distances are introduced in this article. These families generalize the specialization distances presented in [21].

In addition, we prove that some of these distance families are consistent with α -junctions in the sense that two mass functions are closer after performing an α -junction with the same evidence on both of them. This result is mainly explained by the fact that the definition of matrices used in our evidential distances is closely related to α -junctions.

Besides, mass functions are set functions and their domain has a poset structure. It is proved that the sub-family relying on the L^1 matrix norm is compatible with the inclusion partial order if $\alpha > 0$.

Finally, a new aspect of evidential distances is also investigated: the influence

of meta-information on information source truthfulness. The main conclusion to draw is that when two sources lie in the same way, distances relying on L^k matrix norms or on the 1-operator norm are invariant to mass permutations pertaining to the lie if $\alpha = 1$.

There are several perspectives for future works. Concerning theoretical aspects, other polytope metrics in other evidential spaces will be investigated along with their potential new properties. In particular, metrics defined in the Belief space⁵ are likely to be naturally compatible with the inclusion partial order. Concerning practical aspects, we plan to use the distances introduced in this paper for several mass function approximation problems. Indeed, optimization algorithms are highly dependent on the properties of the metrics on which they rely. Finally, it is also intended to study a system containing redundant sensors. As redundant sensors are supposed to lie in the same way, a distance cited in proposition 7 is a relevant candidate to detect a failure among such sensors.

A Proof of Lemma 2

Proof. Let $\mathbf{K}_{i,0}^\cap$ and $\mathbf{K}_{i,0}^\cup$ be the 0-specialization matrix and the 0-generalization matrix of m_i with $i \in \{1; 2\}$. If $\alpha = 0$, equation (3) reduces to:

$$\begin{aligned} m_{1 \cap^0 B}(A) &= \sum_{X \Delta B = A} m_1(X), \\ &= \sum_{X \Delta B = \bar{A}} m_1(X), \\ &= m_{1 \cup^0 B}(\bar{A}). \end{aligned} \quad (32)$$

This is equivalent to $\mathbf{JK}_{i,0}^\cap = \mathbf{K}_{i,0}^\cup$. A similar reasoning as in the proof of proposition 1 gives $d^{\cap,0} = d^{\cup,0}$.

Let \mathbf{K}_i^\cap and \mathbf{K}_i^\cup be the specialization matrix and the generalization matrix of m_i with $i \in \{1; 2\}$. For any A and $B \subseteq \Omega$ we have:

$$\begin{aligned} m_{1 \cup^1 \bar{B}}(A \cup \bar{B}) &= \sum_{X \cup \bar{B} = A \cup \bar{B}} m_1(X), \\ &= \sum_{X \cap B = \bar{A} \cap B} m_1(X). \end{aligned} \quad (33)$$

Let us observe that B is the disjoint union of $B \setminus X$ with $B \setminus \bar{X}$ as well as the disjoint union of $B \setminus A$ with $B \setminus \bar{A}$. Consequently, $B \setminus \bar{X} = B \setminus \bar{A} \Leftrightarrow B \setminus X = B \setminus A$, which gives:

$$\begin{aligned} m_{1 \cup^1 \bar{B}}(A \cup \bar{B}) &= \sum_{X \cap B = A \cap B} m_1(X), \\ &= m_{1 \cap^1 B}(A \cup B). \end{aligned} \quad (34)$$

⁵The Belief space is the subset of \mathbb{R}^N in which belief functions live. Mass functions are in bijective correspondence with belief functions via the Möbius transform.

In other words, matrices \mathbf{K}_i^\cap and \mathbf{K}_i^\cup contain the same elements but their positions are different. Again, a similar reasoning as in the proof of proposition 1 gives $d^{\cap,1} = d^{\cup,1}$. \square

B Proof of Lemma 3

Proof. Let \mathbf{K}_A denote the evidential matrix of any categorical mass function m_A with respect to an α -junction. If $A = B$, the properties are trivially proved. If $A \subsetneq B$, $\exists x \in B$ such that $x \notin A$. Let us now separate the α -conjunctive case and the α -disjunctive case:

- α -conjunctive case: Theorem 13 in [27] gives $\mathbf{K}_A = \prod_{x' \notin A} \mathbf{K}_{\overline{\{x'\}}}$. This allows us to write:

$$\begin{aligned} \|\mathbf{m}_{1 \cap^\alpha A} - \mathbf{m}_{2 \cap^\alpha A}\|_1 &= \|\mathbf{K}_A \mathbf{m}_1 - \mathbf{K}_A \mathbf{m}_2\|_1, \\ &= \|\mathbf{K}_A (\mathbf{m}_1 - \mathbf{m}_2)\|_1, \\ &= \left\| \mathbf{K}_{\overline{\{x\}}} \mathbf{K}_{A \cup \{x\}} (\mathbf{m}_1 - \mathbf{m}_2) \right\|_1. \end{aligned} \quad (35)$$

The very definition of the 1-operator norm for matrix $\mathbf{K}_{\overline{\{x\}}}$ writes as follows:

$$\left\| \mathbf{K}_{\overline{\{x\}}} \right\|_{op1} = \max_{\mathbf{u} \in \mathbb{R}^N} \frac{\left\| \mathbf{K}_{\overline{\{x\}}} \mathbf{u} \right\|_1}{\|\mathbf{u}\|_1}.$$

As $\mathbf{K}_{A \cup \{x\}} (\mathbf{m}_1 - \mathbf{m}_2)$ is a particular vector belonging to \mathbb{R}^N and given that $\left\| \mathbf{K}_{\overline{\{x\}}} \right\|_{op1} = 1$, we have :

$$1 \geq \frac{\left\| \mathbf{K}_{\overline{\{x\}}} \mathbf{K}_{A \cup \{x\}} (\mathbf{m}_1 - \mathbf{m}_2) \right\|_1}{\left\| \mathbf{K}_{A \cup \{x\}} (\mathbf{m}_1 - \mathbf{m}_2) \right\|_1}. \quad (36)$$

Using the above inequality in equation (35) gives:

$$\|\mathbf{m}_{1 \cap^\alpha A} - \mathbf{m}_{2 \cap^\alpha A}\|_1 \leq \left\| \mathbf{m}_{1 \cap (A \cup \{x\})} - \mathbf{m}_{2 \cap (A \cup \{x\})} \right\|_1.$$

The same result can be used by successive inclusions of any other element x' belonging to B while not belonging to A , hence:

$$\|\mathbf{m}_{1 \cap^\alpha A} - \mathbf{m}_{2 \cap^\alpha A}\|_1 \leq \|\mathbf{m}_{1 \cap^\alpha B} - \mathbf{m}_{2 \cap^\alpha B}\|_1.$$

- α -disjunctive case: Theorem 11 in [27] gives $\mathbf{K}_{A \cup \{x\}} = \prod_{x' \in A \cup \{x\}} \mathbf{K}_{\{x'\}}$. This allows us to write:

$$\begin{aligned} \|\mathbf{m}_{1 \cup^\alpha (A \cup \{x\})} - \mathbf{m}_{2 \cup^\alpha (A \cup \{x\})}\|_1 &= \left\| \mathbf{K}_{A \cup \{x\}} \mathbf{m}_1 - \mathbf{K}_{A \cup \{x\}} \mathbf{m}_2 \right\|_1, \\ &= \left\| \mathbf{K}_{A \cup \{x\}} (\mathbf{m}_1 - \mathbf{m}_2) \right\|_1, \\ &= \left\| \mathbf{K}_{\{x\}} \mathbf{K}_A (\mathbf{m}_1 - \mathbf{m}_2) \right\|_1. \end{aligned} \quad (37)$$

The very definition of the 1-operator norm for matrix $\mathbf{K}_{\{x\}}$ writes as follows:

$$\|\mathbf{K}_{\{x\}}\|_{op1} = \max_{\mathbf{u} \in \mathbb{R}^N} \frac{\|\mathbf{K}_{\{x\}}\mathbf{u}\|_1}{\|\mathbf{u}\|_1}.$$

As $\mathbf{K}_A(\mathbf{m}_1 - \mathbf{m}_2)$ is a particular vector belonging to \mathbb{R}^N and given that $\|\mathbf{K}_{\{x\}}\|_{op1} = 1$, one has :

$$1 \geq \frac{\|\mathbf{K}_{\{x\}}\mathbf{K}_A(\mathbf{m}_1 - \mathbf{m}_2)\|_1}{\|\mathbf{K}_A(\mathbf{m}_1 - \mathbf{m}_2)\|_1}. \quad (38)$$

Using the above inequality in equation (37) gives:

$$\|\mathbf{m}_{1 \cup^\alpha (A \cup \{x\})} - \mathbf{m}_{2 \cup^\alpha (A \cup \{x\})}\|_1 \leq \|\mathbf{m}_{1 \cup^\alpha A} - \mathbf{m}_{2 \cup^\alpha A}\|_1.$$

The same result can be used by successive inclusions of any other element x' belonging to B while not belonging to A , hence:

$$\|\mathbf{m}_{1 \cup^\alpha A} - \mathbf{m}_{2 \cup^\alpha A}\|_1 \geq \|\mathbf{m}_{1 \cup^\alpha B} - \mathbf{m}_{2 \cup^\alpha B}\|_1.$$

□

C Proof of corollary 1

Proof. \mathbf{K}_1 and \mathbf{K}_2 are the evidential matrices corresponding to m_1 and m_2 with respect to a given α -junction. Let us use the following definition of the 1-operator matrix norm for matrix $\mathbf{K}_1 - \mathbf{K}_2$:

$$\|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} = \max_{B \subset \Omega} \sum_{A \subset \Omega} |\mathbf{K}_1(A, B) - \mathbf{K}_2(A, B)|. \quad (39)$$

Let us now separate the α -conjunctive case and the α -disjunctive case:

- In case of an α -conjunction, equation (39) also writes:

$$\|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} = \max_{B \subset \Omega} \|\mathbf{m}_{1 \cap^\alpha B} - \mathbf{m}_{2 \cap^\alpha B}\|_1.$$

Using Lemma 3, it is clear that the maximum norm is obtained for the largest set B , *i.e.* when $B = \Omega$, hence:

$$\begin{aligned} \|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} &= \|\mathbf{m}_{1 \cap^\alpha \Omega} - \mathbf{m}_{2 \cap^\alpha \Omega}\|_1, \\ &= \|\mathbf{m}_1 - \mathbf{m}_2\|_1. \end{aligned}$$

- In case of an α -disjunction, equation (39) writes:

$$\|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} = \max_{B \subset \Omega} \|\mathbf{m}_{1 \cup^\alpha B} - \mathbf{m}_{2 \cup^\alpha B}\|_1.$$

Using Lemma 3, it is clear that the maximum norm is obtained for the smallest set B , *i.e.* when $B = \emptyset$, hence:

$$\begin{aligned}\|\mathbf{K}_1 - \mathbf{K}_2\|_{op1} &= \|\mathbf{m}_{1 \cup \alpha \emptyset} - \mathbf{m}_{2 \cup \alpha \emptyset}\|_1, \\ &= \|\mathbf{m}_1 - \mathbf{m}_2\|_1.\end{aligned}$$

□

D Proof of Lemma 4

Proof. Let us first consider that d_1 is defined with respect to an α -conjunctive rule. Let \mathbf{K}_A and \mathbf{K}_B be the evidential matrices corresponding to m_A and m_B with respect to the rule \odot^α and let ρ be a normalization coefficient. By definition of distance d_1 , one can write:

$$\begin{aligned}d_1(m_A, m_B) &= \frac{1}{\rho} \|\mathbf{K}_A - \mathbf{K}_B\|_1, \\ &= \frac{1}{\rho} \sum_{X,Y} |K_A(X, Y) - K_B(X, Y)|.\end{aligned}$$

Given results in [18], it is known that the element $K_A(X, Y)$ of matrix \mathbf{K}_A is non-null if and only if $A \cap Y \subseteq X \subseteq \overline{A\Delta Y}$. Consequently, matrix elements are actually subtracted if $A \cap Y \subseteq X \subseteq \overline{A\Delta Y}$ and $B \cap Y \subseteq X \subseteq \overline{B\Delta Y}$. Let us analyze this situation:

- If X is a superset of both $A \cap Y$ and $B \cap Y$, then X is a superset of $(A \cup B) \cap Y = (A \cap Y) \cup (B \cap Y)$.
- If X is a subset of both $\overline{A\Delta Y}$ and $\overline{B\Delta Y}$, then it is a subset of $\overline{A \cup B \cup Y} \cup (A \cap B \cap Y) = \overline{A\Delta Y} \cap \overline{B\Delta Y}$.

Let $I = A \cap B \cap Y$ and $U = A \cup B \cup Y$. One can write:

$$\begin{aligned}\|\mathbf{K}_A - \mathbf{K}_B\|_1 &= \sum_{\substack{X,Y \\ (A \cup B) \cap Y \subseteq X \subseteq \overline{U \cup I}}} |K_A(X, Y) - K_B(X, Y)| \\ &\quad + \sum_{\substack{X,Y \\ A \cap Y \subseteq X \subseteq \overline{A\Delta Y} \\ B \cap Y \not\subseteq X \text{ or } X \not\subseteq \overline{B\Delta Y}}} K_A(X, Y) + \sum_{\substack{X,Y \\ B \cap Y \subseteq X \subseteq \overline{B\Delta Y} \\ A \cap Y \not\subseteq X \text{ or } X \not\subseteq \overline{A\Delta Y}}} K_B(X, Y), \\ &= \|\mathbf{K}_A\|_1 + \|\mathbf{K}_B\|_1 + \sum_{\substack{X,Y \\ (A \cup B) \cap Y \subseteq X \\ X \subseteq \overline{U \cup I}}} G(X, Y),\end{aligned}$$

with \mathbf{G} a matrix whose elements are such that $G(X, Y) = |K_A(X, Y) - K_B(X, Y)| - K_A(X, Y) - K_B(X, Y)$. Given that the L^1 norm of any evidential matrix is N

and that evidential matrices are positive, we can also write:

$$\|\mathbf{K}_A - \mathbf{K}_B\|_1 = 2N - 2 \sum_{\substack{X, Y \\ (A \cup B) \cap Y \subseteq X \\ X \subseteq \overline{U \cup I}}} K_{A \wedge B}(X, Y),$$

with $\mathbf{K}_{A \wedge B}$ the entry-wise minimum of matrices \mathbf{K}_A and \mathbf{K}_B . It can be proved that $(A \cup B) \cap Y \subseteq \overline{U \cup I} \Leftrightarrow Y \subseteq \overline{A \Delta B}$, which gives:

$$\|\mathbf{K}_A - \mathbf{K}_B\|_1 = 2N - 2 \sum_{\substack{Y \\ Y \subseteq \overline{A \Delta B}}} \sum_{\substack{X \\ (A \cup B) \cap Y \subseteq X \\ X \subseteq \overline{U \cup I}}} K_{A \wedge B}(X, Y).$$

Given that $Y \subseteq \overline{A \Delta B}$, we have $(A \cup B) \cap Y = A \cap B \cap Y$. Consequently, it appears that all subsets X are the union of I with a given subset of \overline{U} . The second sum can thus be re-indexed as follows:

$$\|\mathbf{K}_A - \mathbf{K}_B\|_1 = 2N - 2 \sum_{\substack{Y \\ Y \subseteq \overline{A \Delta B}}} \sum_{\substack{X \\ X \subseteq \overline{U}}} K_{A \wedge B}(X \cup I, Y).$$

From proposition 1 in [18], we have that for any set C such that $C \cap X = \emptyset$:
 $K_{A \wedge B}(X \cup C, Y)$

$$\begin{aligned} &= \min \left\{ \frac{\alpha^{|\overline{A \Delta Y}|}}{\overline{\alpha}^{|A \cap Y|}} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|X \cup C|}; \frac{\alpha^{|\overline{B \Delta Y}|}}{\overline{\alpha}^{|B \cap Y|}} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|X \cup C|} \right\}, \\ &= \min \left\{ \frac{\alpha^{|\overline{A \Delta Y}|}}{\overline{\alpha}^{|A \cap Y|}}; \frac{\alpha^{|\overline{B \Delta Y}|}}{\overline{\alpha}^{|B \cap Y|}} \right\} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|X \cup C|}, \\ &= \min \left\{ \frac{\alpha^{|\overline{A \Delta Y}|}}{\overline{\alpha}^{|A \cap Y|}}; \frac{\alpha^{|\overline{B \Delta Y}|}}{\overline{\alpha}^{|B \cap Y|}} \right\} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|X| + |C|}. \end{aligned}$$

The sum over subsets X can thus be explicitly computed:

$$\sum_{\substack{X \\ X \subseteq \overline{U}}} K_{A \wedge B}(X \cup I, Y)$$

$$\begin{aligned} &= \min \left\{ \frac{\alpha^{|\overline{A \Delta Y}|}}{\overline{\alpha}^{|A \cap Y|}}; \frac{\alpha^{|\overline{B \Delta Y}|}}{\overline{\alpha}^{|B \cap Y|}} \right\} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|I|} \sum_{\substack{X \\ X \subseteq \overline{U}}} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|X|}, \\ &= \min \left\{ \frac{\alpha^{|\overline{A \Delta Y}|}}{\overline{\alpha}^{|A \cap Y|}}; \frac{\alpha^{|\overline{B \Delta Y}|}}{\overline{\alpha}^{|B \cap Y|}} \right\} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|I|} \sum_{i=1}^{|\overline{U}|} \binom{|\overline{U}|}{i} \left(\frac{\overline{\alpha}}{\alpha} \right)^i, \\ &= \min \left\{ \frac{\alpha^{|\overline{A \Delta Y}|}}{\overline{\alpha}^{|A \cap Y|}}; \frac{\alpha^{|\overline{B \Delta Y}|}}{\overline{\alpha}^{|B \cap Y|}} \right\} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|I|} \left(\frac{\overline{\alpha}}{\alpha} + 1 \right)^{|\overline{U}|}, \\ &= \min \left\{ \frac{\alpha^{|\overline{A \Delta Y}|}}{\overline{\alpha}^{|A \cap Y|}}; \frac{\alpha^{|\overline{B \Delta Y}|}}{\overline{\alpha}^{|B \cap Y|}} \right\} \left(\frac{\overline{\alpha}}{\alpha} \right)^{|I|} \alpha^{-|\overline{U}|}. \end{aligned}$$

Besides, since $Y \subseteq \overline{A\Delta B}$, Y is the union of two disjoint subsets $W \subseteq A \cap B$ and $Z \subseteq \overline{A \cup B}$. In particular, we have:

- $|\overline{A\Delta Y}| = |\overline{A\Delta(W \cup Z)}| = |(\overline{A \cup Z}) \setminus W| = n - |A| - |Z| + |W|$ and likewise $|\overline{B\Delta Y}| = n - |B| - |Z| + |W|$,
- $|A \cap Y| = |A \cap (W \cup Z)| = |W|$ and likewise $|B \cap Y| = |W|$,
- $|I| = |W|$,
- $|\overline{U}| = |\overline{A \cup B \cup Z}| = n - |A \cup B| + |Z|$.

This decomposition of Y allows us to write:

$$\begin{aligned} & \min \left\{ \frac{\alpha^{|\overline{A\Delta Y}|}}{\overline{\alpha}^{|A \cap Y|}}; \frac{\alpha^{|\overline{B\Delta Y}|}}{\overline{\alpha}^{|B \cap Y|}} \right\} \\ &= \min \left\{ \frac{\alpha^{n-|A|-|Z|+|W|}}{\overline{\alpha}^{|W|}}; \frac{\alpha^{n-|B|-|Z|+|W|}}{\overline{\alpha}^{|W|}} \right\}, \\ &= \min \left\{ \alpha^{-|A|}; \alpha^{-|B|} \right\} \frac{\alpha^{n-|Z|+|W|}}{\overline{\alpha}^{|W|}}, \\ &= \frac{\alpha^{n-|Z|+|W|-\min\{|A|;|B|\}}}{\overline{\alpha}^{|W|}}. \end{aligned}$$

Going back to the computation of $\|\mathbf{K}_A - \mathbf{K}_B\|_1$, simplifications for the terms at the power of $|W|$ and $|Z|$ are observed, which gives:

$$\begin{aligned} & \|\mathbf{K}_A - \mathbf{K}_B\|_1 \\ &= 2N - 2 \sum_{\substack{W \\ W \subseteq A \cap B}} \sum_{\substack{Z \\ Z \subseteq \overline{A \cup B}}} \alpha^{|A \cup B| - \min\{|A|;|B|\}}, \\ &= 2N - 2\alpha^{|A \cup B| - \min\{|A|;|B|\}} 2^{|A \cap B|} 2^{|\overline{A \cup B}|}, \\ &= 2N \left(1 - \alpha^{|A \cup B| - \min\{|A|;|B|\}} 2^{|A \cap B| - |A \cup B|} \right), \\ &= 2N \left(1 - \frac{\alpha^{\max\{|A \setminus B|; |B \setminus A|\}}}{2^{|A \Delta B|}} \right). \end{aligned}$$

Let us now investigate the α -disjunctive case. For the remainder of this proof, we thus distinguish $d_{1,\cap}$ and $d_{1,\cup}$. From proposition 1, one can write:

$$\begin{aligned} d_{1,\cup}(m_A, m_B) &= d_{1,\cap}(\overline{m}_A, \overline{m}_B), \\ &= d_{1,\cap}(m_{\overline{A}}, m_{\overline{B}}), \\ &= \frac{2N}{\rho} \left(1 - \frac{\alpha^{\max\{|\overline{A} \setminus \overline{B}|; |\overline{B} \setminus \overline{A}|\}}}{2^{|\overline{A \Delta B}|}} \right), \\ &= \frac{2N}{\rho} \left(1 - \frac{\alpha^{\max\{|B \setminus A|; |A \setminus B|\}}}{2^{|A \Delta B|}} \right), \\ &= d_{1,\cap}(m_A, m_B). \end{aligned}$$

□

Acknowledgment

The authors are indebted to the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper and to pave the way for future works.

References

- [1] Mathias Bauer. Approximation algorithms and decision making in the Dempster-Shafer theory of evidence an empirical study. *International Journal of Approximate Reasoning*, 17(23):217 – 237, 1997.
- [2] Thomas Burger and Sebastien Destercke. How to randomly generate mass functions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, pages 645–673, 2013.
- [3] F. Cuzzolin. *Visions of a generalized probability theory*. Lambert Academic Publishing, London, UK, 2014.
- [4] Fabio Cuzzolin. Geometry of Dempster’s rule of combination. *IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics*, 34(2):961 – 977, 2004.
- [5] Fabio Cuzzolin. Two new bayesian approximations of belief functions based on convex geometry. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, Volume 37(4):993–1008, 2007.
- [6] Fabio Cuzzolin. A geometric approach to the theory of evidence. *IEEE Transactions on Systems, Man and Cybernetics - Part C: Application And Reviews*, 38(4):522 – 534, 2008.
- [7] Fabio Cuzzolin. Lp consonant approximation of belief functions in the mass space. In *Proceedings of the 7th International Symposium on Imprecise Probability: Theories and Applications*, pages 1 – 10, Innsbruck, Austria., 2011.
- [8] Fabio Cuzzolin. On consistent approximations of belief functions in the mass space. In Weiru Liu, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty*, volume 6717 of *Lecture Notes in Computer Science*, pages 287–298. Springer Berlin Heidelberg, 2011.
- [9] A.P. Dempster. Upper and lower probabilities induced by a multiple valued mapping. *Annals of Mathematical Statistics*, 38(2):325 – 339, 1967.
- [10] Javier. Diaz, Maria. Rifqi, and Bernadette. Bouchon-Meunier. A similarity measure between basic belief assignments. In *Proceedings of the 9th International Conference on Information Fusion (FUSION 2006)*, pages 1–6, Florence, Italy, 2006.

- [11] Zied Elouedi, Khaled Mellouli, and Philippe Smets. Assessing sensor reliability for multisensor data fusion within the transferable belief model. *IEEE Transactions on Systems, Man, and Cybernetics. Part B: Cybernetics*, 34(1):782 – 787, 2004.
- [12] Dale Fixen and Ronald Mahler. The modified Dempster-Shafer approach to classification. *IEEE Transactions on Systems, Man and Cybernetics. Part A: Systems and humans*, 27(1):96 – 104, 1997.
- [13] Maurice Fréchet. *Sur quelques points du calcul fonctionnel*. PhD thesis, Faculté des Sciences de Paris, 1906.
- [14] Deqiang Han, Jean Dezert, Chongzhao Han, and Yi Yang. New dissimilarity measure in evidence theory. In *Proceedings of the 14th International Conference on Information Fusion (FUSION 2011)*, pages 1 – 7, Chicago, IL. United States, 2011.
- [15] David Harmanec. Faithful approximations of belief functions. In K.B. Laskey and H. Prade, editors, *Uncertainty in Artificial Intelligence (UAI99) 15*, Stockholm (Sweden), 1999.
- [16] Anne-Laure Jousselme, Dominic Grenier, and Eloi Bossé. A new distance between two bodies of evidence. *Information Fusion*, 2:91–101, 2001.
- [17] Anne-Laure Jousselme and Patrick Maupin. Distances in evidence theory : Comprehensive survey and generalizations. *International Journal of Approximate Reasoning*, 53:118 – 145, 2012.
- [18] John Klein, Mehena Loudahi, Jean-Marc Vannobel, and Olivier Colot. α -junctions of categorical mass functions. In Fabio Cuzzolin, editor, *Belief Functions: Theory and Applications*, volume 8764 of *Lecture Notes in Artificial Intelligence*, pages 1–10. Springer International Publishing, 2014.
- [19] George J. Klir and Mark J. Wierman. *Uncertainty-Based Information: Elements of Generalized Information Theory*. Springer, 1999.
- [20] Weiru Liu. Analysing the degree of conflict among belief functions. *Artificial Intelligence*, 170(11):909–924, 2006.
- [21] Mehena Loudahi, John Klein, Jean-Marc Vannobel, and Olivier Colot. New distances between bodies of evidence based on dempsterian specialization matrices and their consistency with the conjunctive combination rule. *International Journal of Approximate Reasoning*, 55(5):1093 – 1112, 2014.
- [22] W.L. Perry and H.E. Stephanou. Belief function divergence as a classifier. In *Intelligent Control, 1991., Proceedings of the 1991 IEEE International Symposium on*, pages 280–285, Aug 1991.

- [23] Frédéric Pichon. On the α -conjunctions for combining belief functions. In Thierry Denœux and Marie-Hlne Masson, editors, *Belief Functions: Theory and Applications*, volume 164 of *Advances in Intelligent and Soft Computing*, pages 285–292. Springer Berlin Heidelberg, 2012.
- [24] Frédéric Pichon and Thierry Denœux. Interpretation and computation of α -junctions for combining belief functions. In *6th International Symposium on Imprecise Probability: Theories and Applications ISIPTA-09*, pages 1 – 10, Durham, United Kingdom, July 2009.
- [25] Frédéric Pichon, Didier Dubois, and Thierry Denœux. Relevance and truthfulness in information correction and fusion. *International Journal of Approximate Reasoning*, 53(2):159 – 175, 2012. Theory of Belief Functions (BELIEF 2010).
- [26] Glenn Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.
- [27] P. Smets. The alpha-junctions: Combination operators applicable to belief functions. In *First Int. Joint Conference on Qualitative and Quantitative Practical Reasoning (ECSQUARU-FAPR '97)*, volume 1244 of *Lecture Notes in Computer Sciences*, pages 131–153. Springer International Publishing, Bad Honef, Germany, 1997.
- [28] Philippe Smets. The application of the matrix calculus to belief functions. *International Journal of Approximate Reasoning*, 31:1–30, 2002.
- [29] Philippe Smets and Robert Kennes. The transferable belief model. *Artificial Intelligence*, 66(2):191 – 234, 1994.
- [30] H.E. Stephanou and S.-Y. Lu. Measuring consensus effectiveness by a generalized entropy criterion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 10(4):544–554, Jul 1988.
- [31] Zachary Sunberg and Jonathan Rogers. A belief function distance metric for orderable sets. *Information Fusion*, 14:361 – 373, 2013.
- [32] Bjornar Tessem. Approximations for efficient computation in the theory of evidence. *Artificial Intelligence*, 61:315 – 329, 1993.
- [33] Thomas Weiler and Ulrich Bodenhofer. Approximation of belief functions by minimizing euclidean distances. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 11(06):170 –177, 2003.
- [34] Lalla Mariem Zouhal and Thierry Denœux. An evidence-theoretic k-nn rule with parameter optimisation. *IEEE Transactions on Systems, Man and Cybernetics. Part C: Application and reviews*, 28(2):263 – 271, 1998.