Geometry of the set of dominating k-additive belief functions

Fabio Cuzzolin, Thomas Burger and Alessandro Antonucci

Department of Computing and Communication Technologies, Oxford Brookes University, Oxford, UK.

Abstract

In this paper we introduce a novel, simpler form of the polytope of inner Bayesian approximations of a belief function, or “consistent probabilities”. We prove that the set of vertices of this polytope is generated by all possible permutations of elements of the domain, mirroring a similar behavior of outer consonant approximations. An intriguing connection with the behavior of maximal outer consonant approximations is highlighted, and the notion of inner (outer) approximation of a credal set in terms of lower probabilities proposed. Finally, we generalize the main result to the case of k-additive belief functions, belief functions whose focal elements have size at most k. We prove that the set of such objects dominating a given belief function is also a polytope whose vertices are generated by permutations of focal elements of size at most k.

Key words: Theory of evidence, consistent probabilities, inner Bayesian approximations, outer consonant approximations, k-additive belief functions, information ordering, permutation.
PACS:

---

1 Introduction

Uncertainty theory is a composite field in which different but related approaches compete to gain a wider audience in engineering [1] and business [2] applications. Belief [3], probability, and possibility [4] measures can all be adopted to represent uncertainty, although some of them may be more fit to

---

Email address: `Fabio.Cuzzolin@brookes.ac.uk` (Fabio Cuzzolin, Thomas Burger and Alessandro Antonucci).
URL: `http://cms.brookes.ac.uk/staff/FabioCuzzolin/` (Fabio Cuzzolin, Thomas Burger and Alessandro Antonucci).

Preprint submitted to Elsevier Science *8 July 2011*

specific applications and contexts.
Their relation is therefore a popular object of study. More specifically, it is useful in many situations to pose the problem of transforming one uncertainty measure into a measure of a different class. In the case of belief functions, this issue goes under the name of “Bayesian” or “consonant” approximation problem, according to whether we seek to approximate a belief measure with a probability or a possibility.
In particular we may request the approximating probabilities to be more “informative” than the original belief function. There are many ways, however, of measuring the information content of a belief function [5–7]. If we adopt the classical ordering induces by belief values

$$b' \geq b \;\equiv\; b'(A) \geq b(A) \quad \forall A \subseteq \Theta$$

we obtain the set of *inner Bayesian approximations* $\mathcal{P}[b]$ of b, i.e., the probability measures whose values dominates those of b for all events A:

$$\mathcal{P}[b] = \Big\{ p \in \mathcal{P} : b(A) \leq p(A) \; \forall A \subseteq \Theta \Big\}. \tag{1}$$

According to Equation (1) this “credal set” [8] of probability measures is in fact the set of probabilities which admit the original b.f. b as a lower bound. A powerful, although controversial, semantic to belief functions comes from the observation that such probabilities can be obtained by redistributing the basic belief or “mass” of each event of Θ to the elements it contains. Belief functions can indeed be seen as constraints acting on the probability simplex, on which they define the credal set (1).

As Shafer admits in [**?**], there is indeed a sense in which a single belief function can be interpreted as a consistent system of probability bounds. However, the issue with such credal interpretation of belief functions becomes evident when considering two or more belief functions addressing the same question but representing conflicting items of evidence, i.e., when Dempster’s rule is applied to aggregate evidence. In [3,9], Shafer disavowed any probability-bound interpretation, a position later seconded by Dempster [10]. Nevertheless this interpretation has been adopted by many authors, and is at the foundation of Smets’ Transferable Belief Model [11–15], in which decisions are made at the *pignistic* level by applying a pignistic transform [16] to convert the available belief function into a probability distribution which turns out to be the barycenter of the associated credal set.

Geometrically, inner Bayesian approximations or *consistent probabilities* are known to form a *polytope* (the convex closure of a finite number of points) in the space of probability measures, whose center of mass coincides with the pignistic transformation [17]. In robust Bayesian statistics a large literature exists on the study of convex sets of probability distributions [18–22].
This polytope of inner Bayesian approximations, in particular, has been stud-

ied by Ha et al. [23], who proved that $\mathcal{P}[b]$ can be expressed in the probability simplex as the sum of the polytopes associated with all the non-zero mass events ("focal elements") A_i, $i = 1, \cdots, k$ of b, weighted by the corresponding masses, i.e.

$$\mathcal{P}[b] = \sum_{i=1}^{k} m_b(A_i) Cl(p_x, x \in A_i) \tag{2}$$

where $Cl(p_x, x \in A_i)$ denotes the convex closure Cl of the probabilities assigning 1 to a particular element x of A_i.

Let us denote more generally by $\mathcal{P}[\mu]$ the set of probability measures dominated by a given capacity (monotone set function) μ. The problem of characterizing the extreme points of $\mathcal{P}[\mu]$ has been studied in the past in greater or smaller generality. Walley has studied the credal set generated by a coherent upper probability in the context of linear previsions, which correspond to finitely additive probabilities [24].

The result that the set of extreme points of $\mathcal{P}[\mu]$ is the following set of probability measures induced by permutations $\rho = (x_\rho(1), ..., x_\rho(n))$ of the elements of the domain $\Theta = \{x_1, ..., x_n\}$ of μ

$$\begin{aligned} &p_\rho(x_\rho(1)) = \mu(\{x_\rho(1)\}), \\ &p_\rho(x_\rho(i)) = \mu(\{x_\rho(1), ..., x_\rho(i)\}) - \mu(\{x_\rho(1), ..., x_\rho(i-1)\}) \quad \forall i = 2, ..., n \end{aligned}$$

when μ is a 2-alternating [25] capacity was first established by Dempster [26] for 1-alternating capacities, while alternative proofs for 2-monotone capacities have been made among others by Shapley [27], in the context of game theory, and Chateauneuf and Jaffray [28], using their Moebius inversion.

Miranda et al [25] have move forward by extending these results to the case of separable metric spaces. On his side, Wallner [29]) has recently provided a characterization of the the maximum number of extreme points of a credal set induced by a lower probability measure (see Section 4).

The case of the set of k-additive dominating belief functions, however, is so far less understood.

In [30,31] the authors have considered the dominance properties of k-additive belief functions for any type of capacities [32], introduced the polytope of k-additive belief functions dominating another belief function, and provided some results to characterize it. Here we will proceed to a full characterization of the vertices of this polytope.

1.1 Contributions

In this paper we start by studying the structure of the vertices of the set of consistent probabilities, proving that the set of actual vertices of the polytope

$\mathcal{P}[b]$ is indeed quite small, and determined by all possible permutations of elements of the domain:

$$\mathcal{P}[b] = Cl(p^\rho[b] \; \forall \rho)$$

where $p^\rho[b]$ is a probability determined by a permutation $\rho = \{x_{\rho(1)}, ..., x_{\rho(n)}\}$ of the singletons of Θ. This allows us to give a simple alternative proof of the pignistic function being the barycenter of such set, and establish a beautiful symmetry with the (dual) case of *outer consonant approximations* [33,34], i.e. the consonant b.f.s dominated by b :

$$\mathcal{O}[b] = \Big\{ co \in \mathcal{CO} : co(A) \leq b(A) \; \forall A \subseteq \Theta \Big\}, \tag{3}$$

where $\mathcal{CO}$ denotes the collection of all consonant b.f.s, i.e. belief functions whose focal elements are nested [3]. As for each maximal chain of focal elements a vertex of the polytope of outer consonant approximations is also determined by a permutation of singletons, there exists a 1-1 correspondence between actual vertices of $\mathcal{P}[b]$ and $\mathcal{O}[b]$.

An application of these results to the issue of approximating an arbitrary credal set with the one generated by a belief function or a lower probability is also outlined.

In the last part of the paper we move on to generalize the result on the credal set of consistent probabilities to analyze the case of the set of k-additive belief functions dominating the original b.f., i.e. the case in which the dominating belief function is allowed to have focal elements of size up to $k \leq n = |\Theta|$. We prove that the vertices of this set are associated with permutations of focal elements of size up to k in this more general case as well.

1.2 Paper outline

We recall in Section 2 the basic notions of the theory of evidence, and the credal interpretation of belief functions. In Section 3 we show that the actual vertices of the polytope of all probability distributions consistent with b are each associated with a permutation of the elements of the domain (3.2). We provide the formal proof, illustrate this proposition in the case of a ternary frame (3.3), and discuss the issue of whether those vertices are guaranteed to be unique (3.4).

In Section 4 an application of this result to credal set approximation for inference on networks is proposed.

In Section 5) we show how the vertices of the region of outer consonant approximations induced by singleton permutations turn out to be in 1-1 correspondence with the vertices of the polygon of consistent probabilities.

Finally, in Section 6 we prove that the vertices of polytope of dominating k-additive belief functions are also associated with permutations (of focal elements of size up to k).

The line of reasoning and all major proofs are illustrated by running examples.

2 Bayesian and consonant approximations of belief functions

Belief [3], probability, and possibility [4] theory are different but related descriptions of uncertainty, as (at least in the finite setting) both probabilities and possibilities are special cases of belief functions.
If we admit that the ideal knowledge state is represented by a "true", but unknown probability measure (which we cannot estimate precisely because of imprecise measurements, missing data, etcetera) belief measures have in turn a natural interpretation as lower/upper bounds to this unknown true probability.

2.1 Belief measures

Belief functions are mathematical representations of the bodies of evidence we possess on a given decision or estimation problem Q. We assume that the possible answers to Q form a finite set $\Theta = \{x_1, ..., x_n\}$ called "frame of discernment". A *basic probability assignment* (b.p.a.) [3] over Θ is a function $m : 2^\Theta \rightarrow [0,1]$ on its power set $2^\Theta = \{A \subseteq \Theta\}$ such that

$$m(\emptyset) = 0, \quad \sum_{A \subseteq \Theta} m(A) = 1.$$

Subsets of Θ associated with non-zero values of m are called "focal elements". The *belief function* (b.f.) $b : 2^\Theta \rightarrow [0,1]$ associated with a basic probability assignment m_b on Θ is defined as

$$b(A) = \sum_{B \subseteq A} m_b(B). \tag{4}$$

In the theory of evidence a probability is just a special belief function assigning non-zero masses to singletons only (*Bayesian* b.f.): $m_b(A) = 0 \; |A| > 1$.
The *plausibility function* (pl.f.) $pl_b : 2^\Theta \rightarrow [0,1]$, $pl_b(A) \doteq 1 - b(A^c)$ measures instead the amount of evidence *not against* A.

2.2 Consistent probabilities or inner Bayesian approximations

The original semantics of belief functions derive from Dempster's analysis of the effect of multi-valued mappings $\Gamma : \Omega \rightarrow 2^\Theta$, $x \in \Omega \mapsto \Gamma(x) \subseteq \Theta$ on evidence available in the form of a probability distribution on the "top" domain

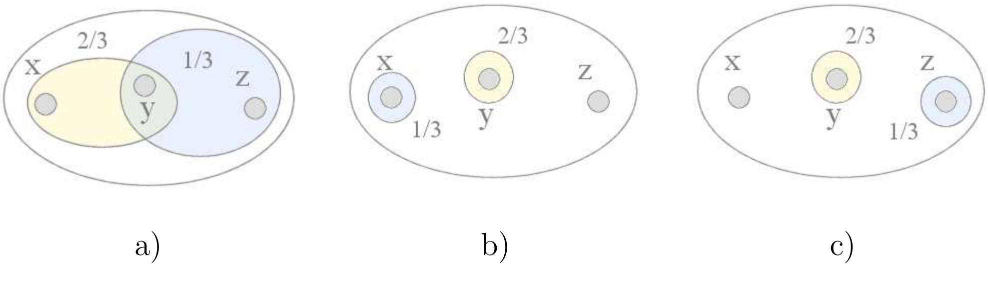
Ω on the "bottom" decision set Θ. As such, belief values are probabilities of events implying other events. In some of his papers [35], however, Dempster himself claimed that the mass $m_b(A)$ associated with a non-singleton event $A \subseteq \Theta$ could be understood as a "floating probability mass" which could not be attached to any particular singleton event $x \in A$ because of the lack of precision of the (multi-valued) operator that quantify our knowledge via the mass function. This has originated a popular but controversial interpretation of belief functions as coherent sets of probabilities determined by sets of lower and upper bounds on their probability values.
According to this interpretation a focal element A of mass $m_b(A)$ can be seen as the indication of the existence of a mass $m_b(A)$ "floating" inside A. The mass assigned to each event $A \subseteq \Theta$ can float freely among its elements $x \in A$.
A probability distribution compatible with b emerges by redistributing the mass of each focal element to its singletons.
This set of Bayesian b.f.s is said "consistent" with b.

2.2.1 Example

To illustrate the notion of probability consistent with a belief function let us consider a little toy example, namely a b.f. b on a frame of cardinality three $\Theta = \{x, y, z\}$ with focal elements (Figure 1-a)

$$m_b(\{x, y\}) = \frac{2}{3}, \quad m_b(\{y, z\}) = \frac{1}{3}. \tag{5}$$

One way of obtaining a probability consistent with b is, for instance, to equally share the mass of $\{x, y\}$ among its elements x and y, while attributing the entire mass of $\{y, z\}$ to y (Figure 1-b). Or, we can assign all the mass of the focal element $\{x, y\}$, and give the mass of $\{y, z\}$ to z only, obtaining the Bayesian belief function of Figure 1-c.



a) b) c)

Fig. 1. a) A simple belief function in the ternary frame. b) and c) two admissible probabilities consistent with a).

2.2.2 *Belief functions as lower bounds*

The set of all and only the (admissible) probabilities obtained as above by re-assigning the mass of each f.e. to its elements is

$$\mathcal{P}[b] \doteq \Big\{ p \in \mathcal{P} : p(A) \geq b(A) \ \forall A \subseteq \Theta \Big\} \qquad (6)$$

i.e., the set of measures whose values dominate that of b on all events A.
A belief function can then be interpreted as a "lower bound" to the set of probabilities it determines.
It is straightforward to check that, for instance, the distribution of Figure 1-b) meets all the lower and upper bounds determined by the belief function (5), for instance:

$$\begin{array}{ll} p(x) & = 1/3 \geq b(x) = 0, \\ p(\{x,y\}) & = 1/3 + 2/3 = 1 \geq b(\{x,y\}) = m_b(\{x,y\}) = 2/3. \end{array}$$

2.2.3 *Consistent belief functions as inner Bayesian approximations*

Consistent probabilities can be seen as the set of all the probabilities *less committed* than the original b.f. b.
The "least commitment principle" [36] postulates that, given a set of b.p.a.s compatible with a number of constraints, the most appropriate mass functions is the "least informative". As pointed out by Denoeux [5], in some sense it plays a role similar to that of maximum entropy in probability theory. However, there are many ways of measuring the information content of a belief function. This is done in practice by defining a partial order in the space of belief functions. Different such orderings were proposed by Yager [6], Dubois and Prade [7], and Denoeux [37]. If we adopt the order relation called *weak inclusion*

$$b \leq b' \equiv b(A) \leq b'(A) \ \ \forall A \subseteq \Theta, \qquad (7)$$

according to which a belief function b' dominates another b.f. b if the belief values of b' are greater than those of b for all events $A \subseteq \Theta$, the consistent probabilities (6) are exactly the group of Bayesian belief functions more committed than b according to (7).
They therefore assume the meaning of *inner Bayesian approximations* of the original belief function b.

3 Vertices of consistent probabilities and element permutations

The region of all inner Bayesian approximations of any given belief function b is a *polytope*, i.e. the convex closure of a finite number of probabilities in the probability simplex.

3.1 The polytope of inner Bayesian approximations

Given an arbitrary belief function b with focal elements $E_1, ..., E_m$, we can define for each choice of m representatives $\{x_1, ..., x_m\}$, $x_i \in E_i \; \forall i$, the (*extremal*) probability measure

$$b_{x_1 \dots x_m} \doteq \sum_{i=1}^{m} m_b(E_i) b_{x_i}. \tag{8}$$

i.e., the Bayesian b.f. obtained by assigning the mass of each focal element E_i to one of its elements $x_i \in E_i$ (recall what said above about the interpretation of focal elements as mass "floating" around in a subset of Θ).

For instance, consider again the belief function (5) of Figure 1-a. If we take as representative x_1 of $E_1 = \{x, y\}$ the element y and as representative x_2 of $E_2 = \{y, z\}$ the element z, we obtain the extremal probability

$$b_{y,z} : m_{b_{y,z}}(y) = m_b(\{x, y\}) = 2/3, \quad m_{b_{y,z}}(z) = m_b(\{y, z\}) = 1/3$$

of Figure 1-c. Note that instead the probability of Figure 1-b, even though is consistent with b, cannot be obtained in this way.

It is well known that $\mathcal{P}[b]$ is indeed the polytope formed by the convex closure of those extremal probabilities [38]:

$$\mathcal{P}[b] = Cl(b_{x_1 \dots x_m}, \{x_1, ..., x_m\} \in E_1 \times \cdots \times E_m). \tag{9}$$

That, though, does not imply that all the points (8) are actual vertices of the simplex $\mathcal{P}[b]$. In other words, several of them may lie on a side of the polytope and be expressed as a convex combination of the others.
In fact, as we show here, the actual vertices of $\mathcal{P}[b]$ can be found in a much smaller set of probabilities, each one associated with a different permutation of the elements of Θ.

3.2 Statement

Theorem 1 *Given a belief function $b : 2^\Theta \to [0,1]$, the simplex $\mathcal{P}[b]$ of the probability measures consistent with b is the polytope*

$$\mathcal{P}[b] = Cl(p^\rho[b] \ \forall \rho)$$

where ρ is any permutation $\{x_{\rho(1)}, ..., x_{\rho(n)}\}$ of the singletons of Θ, and the vertex $p^\rho[b]$ is the Bayesian b.f. such that

$$p^\rho[b](x_{\rho(i)}) = \sum_{A \ni x_\rho(i), A \not\ni x_\rho(j) \ \forall j<i} m_b(A). \tag{10}$$

The probability function (10) attributes to each singleton $x = x_{\rho(i)}$ the mass of all focal elements of b which contain it, but do not contain the elements which precede x in the ordered list $\{x_{\rho(1)}, ..., x_{\rho(n)}\}$ generated by the permutation ρ.

Going back to the binary case $\Theta = \{x, y\}$, it is clear that there exist only two possible permutations of singletons: $\rho^1 : \{x, y\}$, $\rho^2 : \{y, x\}$. They correspond to the following probabilities (Figure 5):

$$p^{x,y}(x) = \sum_{A \supseteq \{x\}} m_b(A) = m_b(x) + m_b(\Theta), \ p^{x,y}(y) = \sum_{A \supseteq \{y\}, A \not\supset \{x\}} m_b(A) = m_b(y);$$
$$p^{y,x}(y) = \sum_{A \supseteq \{y\}} m_b(A) = m_b(y) + m_b(\Theta), \ p^{y,x}(x) = \sum_{A \supseteq \{x\}, A \not\supset \{y\}} m_b(A) = m_b(x).$$

3.3 Proof

We need to prove that:

(1) each probability $p \in \mathcal{P}$ such that $p(A) \geq b(A)$ for all $A \subseteq \Theta$ can be written as a convex combination of the points (10):

$$p = \sum_\rho \alpha_\rho p^\rho[b], \quad \sum_\rho \alpha_\rho = 1, \alpha_\rho \geq 0 \ \forall \rho;$$

(2) vice-versa, each convex combination of the $p^\rho[b]$ satisfies

$$\sum_\rho \alpha_\rho p^\rho[b](A) \geq b(A)$$

for all $A \subseteq \Theta$.

Point (2) is easily proven after we notice that each probability (10) associated with a permutation ρ of elements of Θ is indeed consistent with b, i.e.

$p^\rho[b](A) \geq b(A)\ \forall A$.
Whatever ρ the mass of each of the subsets B of A is attributed by (10) to some element x of A. However, the mass of some other events $B \not\subset A$ is also given to elements of A, so that

$$p^\rho[b](A) = \sum_{x\in A} p^\rho[b](x) \geq \sum_{B\subseteq A} m_b(A) = b(A),$$

i.e., $p^\rho[b]$ is consistent with b whatever the permutation ρ. Therefore

$$\sum_\rho \alpha_\rho p^\rho[b](A) \geq \sum_\rho \alpha_\rho b(A) = b(A) \sum_\rho \alpha_\rho = b(A).$$

As for point (1), we recalled in Section 2.2 that $b'(A) \leq b(A)$ iff m_b is the result of a redistribution of the mass $m_{b'}(A)$ of each f.e. of b' to its subsets. In the case of inner Bayesian approximations the mass of each event has to be redistributed among its elements $x \in A$:

$$m_b(A) \mapsto \alpha_x^A m_b(A) \qquad \forall x \in A, \ \sum_{x\in A} \alpha_x^A = 1. \tag{11}$$

Therefore, under such a redistribution process, for all p such that $b(A) \leq p(A)$ $\forall A$ the mass $p(x)$ of each element $x \in \Theta$ is

$$p(x) = \sum_{A\supseteq\{x\}} m_b(A) \alpha_x^A. \tag{12}$$

To prove point (1) we need therefore to rewrite (12) as a convex combination of the $p^\rho[b](x)$, i.e. we have to enforce

$$p(x) = \sum_\rho \alpha_\rho p^\rho[b](x) = \sum_\rho \alpha_\rho \left(\sum_{A\ni x=x_{\rho(i)}, A\not\ni x_{\rho(j)} \forall j<i} m_b(A) \right)$$

where i is the position of the element x according to the permutation ρ.
For all $A \supseteq \{x\}$ there exists a permutation ρ such that the elements before x in $\{x_{\rho(1)}, ..., x_{\rho(n)}\}$ fall outside A. Hence the above quantity reads as

$$\sum_{A\supseteq\{x\}} m_b(A) \left(\sum_{\rho: x_{\rho(j)} \notin A \forall j<i} \alpha_\rho \right)$$

where again $x = x_{\rho(i)}$.

In summary we need to show that the system of equations

$$\left\{ \alpha_x^A = \sum_{\rho: x_{\rho(j)}\notin A \forall j<i, x=x_{\rho(i)}} \alpha_\rho \quad \forall x \in \Theta, \quad \forall A \supseteq \{x\} \right. \tag{13}$$

has at least one solution $\{\alpha_\rho\}$ such that $\sum_\rho \alpha_\rho = 1$, $\alpha_\rho \geq 0\ \forall \rho$.

3.3.1 Parenthesis: proof in the ternary case

It is useful to first illustrate the existence of a convex solution to (13) in the simple but interesting case of a ternary frame $\Theta = \{x, y, z\}$.
The possible permutations of singletons in this case are six:

$$\begin{array}{lll} \rho^1 = \{x, y, z\}, & \rho^2 = \{x, z, y\}, & \rho^3 = \{y, x, z\}, \\ \rho^4 = \{y, z, x\}, & \rho^5 = \{z, x, y\}, & \rho^6 = \{z, y, x\}. \end{array}$$

The system of equations (13) reads as

$$\left\{ \begin{array}{lllllll}
\alpha_x^{\{x\}} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +\alpha_{\rho^3} & +\alpha_{\rho^4} & +\alpha_{\rho^5} & +\alpha_{\rho^6}; \\
\alpha_x^{\{x,y\}} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +0 & +0 & +\alpha_{\rho^5} & +0; \\
\alpha_x^{\{x,z\}} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +\alpha_{\rho^3} & +0 & +0 & +0; \\
\alpha_x^{\Theta} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +0 & +0 & +0 & +0; \\
\alpha_y^{\{y\}} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +\alpha_{\rho^3} & +\alpha_{\rho^4} & +\alpha_{\rho^5} & +\alpha_{\rho^6}; \\
\alpha_y^{\{x,y\}} & = 0 & +0 & +\alpha_{\rho^3} & +\alpha_{\rho^4} & +0 & +\alpha_{\rho^6}; \\
\alpha_y^{\{y,z\}} & = \alpha_{\rho^1} & +0 & +\alpha_{\rho^3} & +\alpha_{\rho^4} & +0 & +0; \\
\alpha_y^{\Theta} & = 0 & +0 & +\alpha_{\rho^3} & +\alpha_{\rho^4} & +0 & +0; \\
\alpha_z^{\{z\}} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +\alpha_{\rho^3} & +\alpha_{\rho^4} & +\alpha_{\rho^5} & +\alpha_{\rho^6}; \\
\alpha_z^{\{x,z\}} & = 0 & +0 & +0 & +\alpha_{\rho^4} & +\alpha_{\rho^5} & +\alpha_{\rho^6}; \\
\alpha_z^{\{y,z\}} & = 0 & +\alpha_{\rho^2} & +0 & +0 & +\alpha_{\rho^5} & +\alpha_{\rho^6}; \\
\alpha_z^{\Theta} & = 0 & +0 & +0 & +0 & +\alpha_{\rho^5} & +\alpha_{\rho^6}.
\end{array} \right.$$

As by definition $\alpha_x^x = 1$, all equations associated with a singleton generate the normalization constraint: $\alpha_{\rho^1} + \alpha_{\rho^2} + \alpha_{\rho^3} + \alpha_{\rho^4} + \alpha_{\rho^5} + \alpha_{\rho^6} = 1$.
Also, many equations in the above system are actually linearly dependent, such as for instance equations 2 and 6 (associated with $\alpha_x^{\{x,y\}}$ and $\alpha_y^{\{x,y\}}$). This is due to the fact that, by definition (11),

$$\sum_{x \in A} \alpha_x^A = 1. \tag{14}$$

After eliminating the dependencies we get a reduced system

$$\left\{\begin{array}{llllllll}
\alpha_z^{\Theta} & = 0 & +0 & +0 & +0 & +\alpha_{\rho^5} & +\alpha_{\rho^6}; \\
\alpha_x^{\{x,y\}} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +0 & +0 & +\alpha_{\rho^5} & +0; \\
\alpha_y^{\{y,z\}} & = \alpha_{\rho^1} & +0 & +\alpha_{\rho^3} & +\alpha_{\rho^4} & +0 & +0; \\
\alpha_x^{\{x,z\}} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +\alpha_{\rho^3} & +0 & +0 & +0; \\
\alpha_x^{\Theta} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +0 & +0 & +0 & +0; \\
\alpha_x^{\{x\}} & = \alpha_{\rho^1} & +\alpha_{\rho^2} & +\alpha_{\rho^3} & +\alpha_{\rho^4} & +\alpha_{\rho^5} & +\alpha_{\rho^6};
\end{array}\right.$$

which obviously admits as solution

$$\begin{array}{lll}
\alpha_{\rho^6} = \alpha_z^{\Theta}, & \alpha_{\rho^5} = \alpha_x^{\{x,y\}}, & \alpha_{\rho^4} = \alpha_y^{\{y,z\}}, \\
\alpha_{\rho^3} = \alpha_x^{\{x,z\}}, & \alpha_{\rho^2} = \alpha_x^{\Theta}, & \alpha_{\rho^1} = \alpha_x^{\{x\}}
\end{array}$$

which in turns represents a valid convex combination of the $p^{\rho}[b]$.
Of course there are many ways of obtaining a reduced system, and therefore many acceptable convex solutions to (13).

3.3.2 General solution

As we have learned in the ternary case, the normalization constraint is in fact trivially satisfied, as from (13) it follows that when $A = \{x\}$, $x \in \Theta$

$$1 = \alpha_x^x = \sum_{\rho : x_{\rho(j)} \notin \{x\} \forall j < i, x = x_{\rho(i)}} \alpha_\rho = \sum_\rho \alpha_\rho,$$

i.e., $\sum_\rho \alpha_\rho = 1$. Let us denote by $x_{|A|}$ any element representative of A. Due to the normalization constraint (14) the system of equations (13) reduces to

$$\left\{ \alpha_x^A = \sum_{\rho : x_{\rho(j)} \notin A \forall j < i, x = x_{\rho(i)}} \alpha_\rho \quad \forall A \subseteq \Theta, x \neq x_{|A|}. \right. \tag{15}$$

Again $\forall A$ s.t. $|A| = 1$ we get simply the normalization constraint.

To understand the structure of (15) consider some arbitrary ordering of the elements of Θ, $x_1, ..., x_n$. If we take as representative of any event A its last element according to this ordering, we can write (15) as

$$\left\{ \alpha_{x_k}^A = \sum_{\rho : x_{\rho(j)} \notin A \forall j < i, x_k = x_{\rho(i)}} \alpha_\rho, \quad \begin{array}{l} A \supseteq \{x_k\} \\ A \not\subset \{x_1, ..., x_k\} \end{array} \right. \tag{16}$$

each block associated with x_k, $k = 1, ..., n-1$. The number of equations in each block k for a frame Θ of size $|\Theta| = n$ is

$$\begin{aligned} & |\{A \subseteq \Theta : A \supseteq \{x_k\}, A \not\subset \{x_1, ..., x_k\}\}| = \\ & = |\{A \subseteq \Theta : A \supseteq \{x_k\}, A \cap \{x_1, ..., x_k\}^c \neq \emptyset\}| = 2^{i-1}(2^{n-i} - 1) = 2^{n-1} - 2^{i-1}. \end{aligned}$$

Now, all the equations of each block k involve (amongst others) the α_ρ related to permutations ρ which put x_k in the first position: $x_k = x_{\rho(1)}$, as it obviously has no predecessors so that there is no $j < i$ in the subscript of the sum in (15) or (16).
The number of such permutations is clearly $(n-1)!$ (the number of possible orderings of the $n-1$ successors of x_k).

But for $n > 4$ we have that $(n-1)! \geq 2^{n-1} - 2^{i-1}$: Each block has less equations than the number of permutations associated with variables α_ρ which appear in all the equations of the block. Therefore we can assign the first term of each equation of the block to one of those variables: $\alpha^A_{x_k} = \alpha_\rho$ for some ρ which puts x_k in the first position, this for all $A : A \supseteq \{x_k\}, A \not\subset \{x_1, ..., x_k\}$ (all equations in the block).
Variables associated with the remaining permutations can be set to zero. This yields a convex solution to (15), and therefore to the original system (13).

If $n = 3$ we have seen that a solution also exists. For $n = 2$ the solution is trivial. If $n = 4$ the condition $(n-1)! \geq 2^{n-1} - 2^{i-1}$ still holds for all blocks but the first one, for which the number of equations is $2^{n-1} - 2^{i-1} = 7$ while the number of variables in common is $(n-1)! = 6$. But it suffices to use the normalization constraint to replace the equation for Θ in the block x_1 (which is in excess) with an equation for Θ in the block x_3 and obtain an equivalent system of equations which meets the desired property. $\square$

3.4 Uniqueness

We may wonder whether all the extremal points (10) generated by distinct permutations of singletons are guaranteed to be distinct. The answer is negative. Consider a belief function

$$\begin{array}{rrr} m_b(x) = 0.2, & m_b(y) = 0.1, & m_b(z) = 0.3, \\ m_b(\{x,y\}) = 0.1, & m_b(\{y,z\}) = 0.2, & m_b(\Theta) = 0.1 \end{array} \tag{17}$$

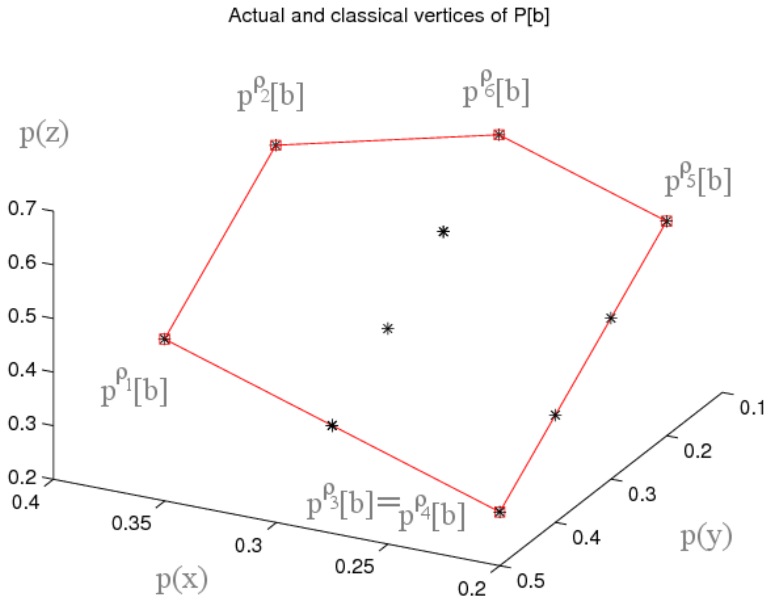


Fig. 2. The actual number of vertices (10), (18) of $\mathcal{P}[b]$ (red squares) is much smaller than the number of candidate points (8) of the classical expression. Here they are plotted as black stars for the belief function (17) of the example. Some of them even fall inside the polytope.

defined on a ternary frame $\Theta = \{x, y, z\}$. In this case there are six possible element permutations. Therefore by Theorem 1 $\mathcal{P}[b]$ has as vertices

$$\begin{array}{lll} \rho^1 = \{x,y,z\} : p^{\rho^1}[b](x) = .4, & p^{\rho^1}[b](y) = .3, & p^{\rho^1}[b](z) = .3; \\ \rho^2 = \{x,z,y\} : p^{\rho^2}[b](x) = .4, & p^{\rho^2}[b](y) = .1, & p^{\rho^2}[b](z) = .5; \\ \rho^3 = \{y,x,z\} : p^{\rho^3}[b](x) = .2, & p^{\rho^3}[b](y) = .5, & p^{\rho^3}[b](z) = .3; \\ \rho^4 = \{y,z,x\} : p^{\rho^4}[b](x) = .2, & p^{\rho^4}[b](y) = .5, & p^{\rho^4}[b](z) = .3; \\ \rho^5 = \{z,x,y\} : p^{\rho^5}[b](x) = .3, & p^{\rho^5}[b](y) = .1, & p^{\rho^5}[b](z) = .6; \\ \rho^6 = \{z,y,x\} : p^{\rho^6}[b](x) = .2, & p^{\rho^6}[b](y) = .2, & p^{\rho^6}[b](z) = .6; \end{array} \tag{18}$$

and we can notice that the permutations $\rho^3 = \{y, x, z\}$ and $\rho^4 = \{y, z, x\}$ yield the same function: $p^{\rho^3}[b] = p^{\rho^4}[b]$.

Notice also that according to the classical expression (9) of $\mathcal{P}[b]$ there are many more (candidate) vertices (8), namely

$$\prod_{A \subseteq \Theta : m_b(A) \neq 0} |A|.$$

Many of those fall on some side of $\mathcal{P}[b]$, or even its interior (see Figure 2).

4 Credal set approximation of credal sets via belief functions

The form of the credal set associated with a belief function, and in particular its limited number of vertices, can be attractive in contexts in which we need to work with credal sets in an efficient way.

A (partial) characterization of a credal set can be obtained by considering its *lower probabilities*, i.e., the infima (over all the probability mass functions in the set) of the probabilities assigned to the elements of the power set. These bounds correspond to a number of constraints satisfied by the original credal set. Yet, this is only a partial characterization, as the set of probability mass functions consistent with these constraints is in general a proper superset of the original credal set. In [39] such an "outer approximation" of credal sets was studied for its potential applications to inference on *credal networks* [40]. Indeed, in order to achieve this approximation only the lower probabilities of a credal are needed, while an explicit enumeration of its extreme points is not necessary. This makes the approximation particularly suited for credal networks, which are a generalization of Bayesian networks based on CSs. In fact, most of the algorithms for credal networks updating only return the lower posterior probabilities. In [39] we showed that the outer approximation of this posterior credal set, as returned by the proposed outer transformation, can be computed by means of these standard algorithms. This makes it possible to adopt more refined criteria for making decisions on a credal network, without the need of newer algorithms.
Although the outer approximation can eventually lead to over-cautious decisions, numerical simulations show that this happens only in a minority of cases.
Consider as an example the following credal set K in the probability simplex of a domain $\Theta = \{x, y, z\}$ with just three elements:

$$K = Cl \left\{ \begin{bmatrix} .90 \\ .05 \\ .05 \end{bmatrix}, \begin{bmatrix} .10 \\ .40 \\ .50 \end{bmatrix}, \begin{bmatrix} .20 \\ .20 \\ .60 \end{bmatrix}, \begin{bmatrix} .20 \\ .70 \\ .10 \end{bmatrix}, \begin{bmatrix} .80 \\ .05 \\ .15 \end{bmatrix}, \begin{bmatrix} .45 \\ .25 \\ .30 \end{bmatrix}, \begin{bmatrix} .05 \\ .80 \\ .15 \end{bmatrix} \right\}, \qquad (19)$$

where probability mass functions are denoted as vertical arrays. It is easy to verify that none of these seven probability mass functions is a convex combination of the remaining six. Hence this set has $7 > 3! = 6$ vertices, and cannot correspond to a belief function or a lower probability measure.

Now, consider the "outer" approximation defined as follows:

Definition 1 *Given a credal set K on Θ, consider its extreme points $\mathcal{V}(K)$.*

Then, for each $A \subseteq \Theta$*, compute the* lower probability*:*[1]

$$\underline{P}_K(A) := \min_{p \in \mathcal{V}(K)} \sum_{x \in A} p(x). \tag{20}$$

We call the credal set $\tilde{K}$ *associated with the lower probability operator in (20) as follows:*

$$\tilde{K} := \left\{ p \in \overline{K} : \sum_{x \in A} p(x) \geq \underline{P}_K(A), \forall A \in 2^\Theta \right\} \tag{21}$$

(where $\overline{K}$ *is the convex closure of* K*) the* outer approximation *of* K*.*

When applying this outer approximation to the credal set (19) we obtain, according to (21), the following lower probability values

$$\begin{cases} P(\{x\}) \leq .05 \\ P(\{y\}) \leq .05 \\ P(\{z\}) \leq .05 \\ P(\{x, y\}) \leq .40 \\ P(\{x, z\}) \leq .20 \\ P(\{y, z\}) \leq .10. \end{cases}$$

from which the outer approximation is:

$$\tilde{K} = Cl \left\{ \begin{bmatrix} .05 \\ .35 \\ .60 \end{bmatrix}, \begin{bmatrix} .05 \\ .80 \\ .15 \end{bmatrix}, \begin{bmatrix} .15 \\ .80 \\ .05 \end{bmatrix}, \begin{bmatrix} .35 \\ .05 \\ .60 \end{bmatrix}, \begin{bmatrix} .90 \\ .05 \\ .05 \end{bmatrix} \right\}. \tag{22}$$

Figure 4 depicts the polytopes associated to $K(X)$ and $\tilde{K}(X)$ on the same probability simplex, clearly illustrating the reason for calling $\tilde{K}(X)$ the "outer approximation" of K.

It can be proven that:

Theorem 2 *Consider a credal set* K *on* Θ*. Let* $\underline{P}_K$ *denote the corresponding lower probability operator as in (20), and* $\tilde{K}$ *the output of the outer transformation of Definition 1. Then:*

- $K \subseteq \tilde{K}$*;*

---

[1] The minimum in (20) is the same we obtain by minimizing over the whole credal set K [41].

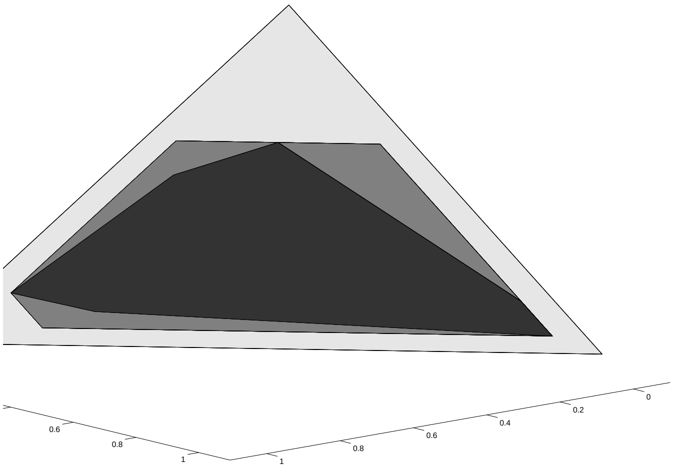


Fig. 3. The credal set K of the example (19) (dark gray), and the output $\tilde{K}$ of the outer transformation (21) of Equation (22) (medium gray) in the probability simplex (light gray) associated with a domain $\Theta = \{x, y, z\}$ with three elements.

- $K = \tilde{K}$ *if and only if a lower probability operator* $\underline{P}'$ *such that* $K_{\underline{P}'} = K$ *exists;*
- *A lower probability operator* $\underline{P}' \neq \underline{P}_K$ *such that* $K \subseteq K_{\underline{P}'} \subseteq \tilde{K}$ *cannot exist.*

i.e., $\tilde{K}$ is the minimal outer approximation of K induced by a lower probability.

Different techniques can be adopted to evaluate the quality of the proposed outer approximation. As an example, a geometrical approach would consist in comparing the area of the polytopes associate to the two credal sets. Yet, as a credal set is basically of model of uncertain knowledge to be used to make decisions, it seems more reasonable to compare decisions based on the two sets [39]. A similar *inner approximation* can also be introduced. The study of such approximations, the appropriate dissimilarity measures between credal sets, and their application to inference is an exciting open field of research.

5 Bayesian and consonant approximations: a symmetry

Besides including finite probabilities are a special case, belief measures also generalize finite *possibility* [42] measures, i.e. functions $Pos : 2^\Theta \rightarrow [0, 1]$ on Θ such that $Pos(\bigcup_i A_i) = \sup_i Pos(A_i)$ for any family of sets $\{A_i, i \in I\}$ (where I is an arbitrary set index).
More precisely, a b.f. is *consonant* (co.b.f.) when its focal elements $\{E_i, i = 1, ..., m\}$ are nested: $E_1 \subset E_2 \subset ... \subset E_m$. As a matter of fact it can be proven

that [4,43] the plausibility function pl_b associated with a belief function b on a domain Θ is a possibility measure iff b is consonant.

As possibility measures form a subclass of belief functions we can pose the problem of approximate a belief function with a possibility (or equivalently with a consonant b.f.) in perfect analogy to the case of Bayesian approximation. In particular, "outer consonant approximations" form a dual couple with inner Bayesian approximations or consistent probabilities.

5.1 Outer consonant approximations

We call *outer consonant approximations* of a belief function b [34] all the co.b.f.s which are *less committed* than the original belief function b:

$$\mathcal{O}[b] = \Big\{ co \in \mathcal{CO} : co(A) \leq b(A) \ \forall A \subseteq \Theta \Big\}. \tag{23}$$

Here $\mathcal{CO}$ denotes the set of all consonant b.f.s.
According to the interpretation of the weak inclusion relation discussed in Section 2.2.3, $b' \leq b$ is equivalent to say that b' is obtained by letting the mass of each focal element A of b float to one or more events containing A: $B \supseteq A$. If b' is also consonant, its focal elements have to form a chain $E_1 \subset \cdots \subset E_n$, $|E_i| = i$. An outer consonant approximation of b is then obtained by letting the mass of each f.e. be re-distributed to one or more elements of the chain.

5.1.1 Example

As an example, an outer consonant approximation of the belief function (5) of Figure 1 can be obtained by re-assigning the mass 2/3 of $A = \{x, y\}$ one half (1/3) to $\{x, y\}$ itself and one half (1/3) to $\Theta \supset \{x, y\}$, and the mass of $A = \{y, z\}$ to $\Theta \supset \{y, z\}$ also. What we get is a consonant b.f. with focal elements $\{x, y\} \subset \Theta$ and b.p.a.

$$m'(\{x, y\}) = 2/3, \quad m'(\Theta) = 1/3.$$

5.1.2 Outer consonant approximations generated by permutations

In particular, with the purpose of finding outer approximations which are minimal with respect to the weak inclusion relation (7), Dubois and Prade have introduced two different families of approximations [33].
In particular a first group of outer consonant approximations can be obtained by considering all permutations ρ of the elements $\{x_1, ..., x_n\}$ of the frame of

discernment Θ: $\{x_{\rho(1)}, ..., x_{\rho(n)}\}$. A family of nested sets can be then built

$$\Big\{S_1^\rho = \{x_{\rho(1)}\}, S_2^\rho = \{x_{\rho(1)}, x_{\rho(2)}\}, \cdots, S_n^\rho = \{x_{\rho(1)}, ..., x_{\rho(n)}\}\Big\} \tag{24}$$

so that a new consonant belief function co^ρ can be defined with b.p.a.

$$m_{co^\rho}(S_j^\rho) = \sum_{i:E_i \subseteq S_j^\rho, E_i \not\subset S_{j-1}^\rho} m_b(E_i). \tag{25}$$

S_j^ρ concentrates all the mass of the focal elements E_i of b included in S_j^ρ but not in S_{j-1}^ρ.

5.1.3 Example

Let us consider again the belief function (5) of the example of Section 2.2.1. A possible permutation of the singletons of Θ is, for instance,

$$\rho = \{x_{\rho(1)}, x_{\rho(2)}, x_{\rho(3)}\} = \{y, z, x\}.$$

This permutation generates the following list of nested sets (24):

$$\Big\{S_1^\rho = \{x_{\rho(1)}\} = \{y\},\ S_2^\rho = \{x_{\rho(1)}, x_{\rho(2)}\} = \{y, z\},$$
$$S_3^\rho = \{x_{\rho(1)}, x_{\rho(2)}, x_{\rho(3)}\} = \{x, y, z\}\Big\}.$$

By Equation (25) we assign to $S_1^\rho = \{y\}$ the mass of all focal elements (5) of b included in $\{y\}$: there are none, so that $m_{co^\rho}(\{y\}) = 0$. To $S_2^\rho = \{y, z\}$ we assign the mass of all f.e.s inside $\{y, z\}$ not contained in $\{y\}$, i.e. the mass 1/3 of $\{y, z\}$ itself. Finally, $S_3^\rho = \{x, y, z\}$ is assigned the mass of all f.e.s which are subsets of $\{x, y, z\}$, but not of $S_2^\rho = \{y, z\}$, namely the mass 2/3 of $\{x, y\}$ (see Figure 4).

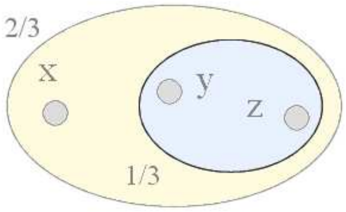


Fig. 4. The outer consonant approximation of (5) generated by the permutation of singletons $\rho = \{y, z, x\}$.

5.2 Binary case

In the binary case (Figure 5) a compelling symmetry emerges between $\mathcal{O}[b]$ and $\mathcal{P}[b]$. Given the definition of weak inclusion (7) it is straightforward to

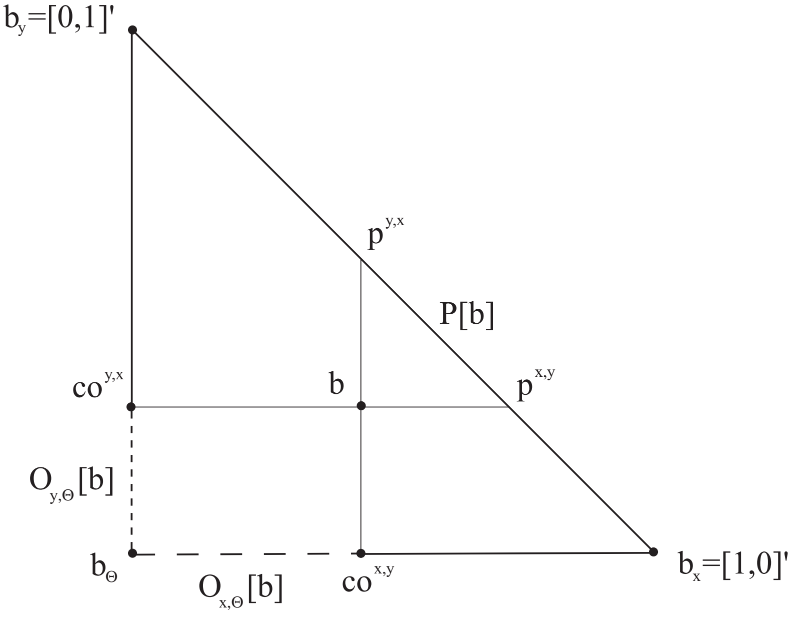


Fig. 5. Geometry of outer consonant approximations and inner Bayesian approximations in the binary case.

recognize that (as the abscissa measures the degree of belief $b(x)$ of x, and the ordinate the d.o.f. $b(y)$) the set of inner Bayesian approximations and the set of outer consonant approximations form respectively a segment $\mathcal{P}[b]$ delimited by a pair of probabilities, and the union of two segments $O_{x,\Theta}[b]$ and $O_{y,\Theta}[b]$.

From (Figure 5) we can remark the existence of an apparent bijection between vertices of $\mathcal{O}[b]$ and $\mathcal{P}[b]$. The vertex $p^{y,x}$ of the interval of consistent probabilities has the same belief value on x as the vertex $co^{x,y}$ of outer consonant approximation associated with the opposite permutation $\{x, y\}$.
In fact we can prove that the outer consonant approximations (25) generated by permutations of singletons form a subset of the vertices of $\mathcal{O}[b]$ [44] even in the general case. Furthermore, the correspondence between vertices of $\mathcal{O}[b]$ and $\mathcal{P}[b]$ generated by permutations of singletons holds in the general case.

5.3 1-1 correspondence

Indeed, the family of outer consonant approximations (25) generated by permutations of singletons is linked by a very elegant geometric duality to the vertices (10) of the polytope of consistent probabilities $\mathcal{P}[b]$.
Let us go back to the example of the ternary frame $\Theta = \{x, y, z\}$. Given for instance the trivial permutation $\rho = \{x, y, z\}$ the vertex co^ρ has focal elements

$\{x\}$, $\{x, y\}$, and $\{x, y, z\}$ and b.p.a.

$$\begin{aligned} m_{co^\rho}(\{x\}) &= m_b(x); \\ m_{co^\rho}(\{x,y\}) &= \sum_{A\subseteq\{x,y\}, A\not\subset\{x\}} m_b(A) = m_b(\{y\}) + m_b(\{x,y\}); \\ m_{co^\rho}(\Theta) &= \sum_{A\subseteq\Theta, A\not\subset\{x,y\}} m_b(A) \\ &= m_b(\{z\}) + m_b(\{x,z\}) + m_b(\{y,z\}) + m_b(\Theta). \end{aligned}$$

Consider, instead, the reverse permutation $\bar\rho = \{z, y, x\}$. The corresponding vertex $p^{\bar\rho}$ of $\mathcal{P}[b]$ has b.p.a.

$$\begin{aligned} p^{\bar\rho}(z) &= \sum_{A\supseteq\{z\}} m_b(A) = m_b(\{z\}) + m_b(\{x,z\}) + m_b(\{y,z\}) + m_b(\Theta); \\ p^{\bar\rho}(y) &= \sum_{A\supseteq\{y\}, A\not\supset\{z\}} m_b(A) = m_b(\{y\}) + m_b(\{x,y\}); \\ p^{\bar\rho}(x) &= \sum_{A\supseteq\{x\}, A\not\supset\{y\},\{z\}} m_b(A) = m_b(x) \end{aligned}$$

i.e., the b.p.a.s of co^ρ and $p^{\bar\rho}$ *coincide*. This is true in the general case.

Theorem 3 *Consider a belief function $b : 2^\Theta \to [0, 1]$. There exists a 1-1 correspondence between the vertices co^ρ (25) of the region of outer consonant approximations of b generated by permutations ρ of the singletons, and the vertices p^ρ (10) of the polytope $\mathcal{P}[b]$ of probabilities consistent with b (all of which are also associated with permutations of singletons).*
Namely, given a permutation $\rho = \{x_{\rho(1)}, ..., x_{\rho(n)}\}$ of the elements of Θ and its inverse $\bar\rho = \{x_{\bar\rho(1)}, ..., x_{\bar\rho(n)}\} = \{x_{\rho(n)}, ..., x_{\rho(1)}\}$ we have that

$$p^{\bar\rho}(x_{\bar\rho(i)}) = m_{co^\rho}(\{x_{\rho(1)}, ..., x_{\rho(n-i+1)}\}) \tag{26}$$

i.e. their b.p.a.s on $\{S^\rho_n, ..., S^\rho_1\}$, $\{x_{\bar\rho(1)}, ..., x_{\bar\rho(n)}\}$ (respectively) coincide as vectors of length n.

Proof. It suffices to show that, as $\bar\rho(i) = \rho(n - i + 1)$,

$$\begin{aligned} p^{\bar\rho}(x_{\bar\rho(i)}) &= \sum_{A\ni x_{\bar\rho(i)}, A\not\ni x_{\bar\rho(j)} \forall j<i} m_b(A) \\ &= \sum_{A\ni x_{\rho(n-i+1)}, A\not\ni x_{\rho(j)} \forall j>n-i+1} m_b(A) \\ &= \sum_{\substack{A \subseteq \{x_{\rho(1)}, ..., x_{\rho(n-i+1)}\}, \\ A \not\subset \{x_{\rho(1)}, ..., x_{\rho(n-i)}\}}} m_b(A) = m_{co^\rho}(\{x_{\rho(1)}, ..., x_{\rho(n-i+1)}\}) \\ &= m_{co^\rho}(S^\rho_{n-i+1}). \end{aligned}$$

□

6 The polytope of k-additive dominating belief functions

6.1 k-additive belief functions

A *k-additive belief function* b on Θ (with $k \in \mathbb{N}$ and $k \leq |\Theta|$) is a belief function such that its mass function m has no focal elements of cardinality $> k$. Clearly:

(1) a Bayesian belief function on Θ is a 1-additive belief function;
(2) for any belief function b on Θ there is a $k \in \{1, ..., |\Theta|\}$ such that b is k-additive.

k-additive belief functions were originally introduced in [45]. Their interest is manyfold. Firstly, the number of degrees of freedom (the number of possible focal elements) of a k-additive belief function is:

$$|\mathcal{P}_k(\Theta)| = \sum_{i=1}^{i=k} \binom{|\Theta|}{i}$$

whereas the number of degrees of freedom of a generic belief function is:

$$|\mathcal{P}(\Theta)| = \underbrace{\sum_{i=1}^{i=\Theta} \binom{|\Theta|}{i}}_{2^{|\Theta|}} = |\mathcal{P}_k(\Theta)| + \underbrace{\sum_{i=k+1}^{i=|\Theta|} \binom{|\Theta|}{i}}_{>0 \text{ if } |\Theta|>k}.$$

making k-additive b.f.s a more compact representation of knowledge. As computation cost is one of the major drawbacks of the theory of belief functions, most frequently criticized in the literature, the use of k-additive representations is of real interest from a computational point of view.

As mentioned in [45], another advantage of k-additive belief functions is that they are easier to handle from a perceptive point of view. Humans find it rather difficult to attach meaning to focal elements of larger cardinality. Hence, limiting the focal elements to subsets of Θ with bounded cardinality is a sensible way of ensuring that the corresponding mathematical representation of knowledge is intuitive and easy to handle.

Finally, as k-additive belief functions are somehow intermediate objects between belief functions and probabilities (in terms of their degrees of freedom), they constitute an interesting trade-off between the full expressive power of belief functions and the simplicity of interpretation of probability measures (as 1-additive belief functions).

According to Theorem 1, the polytope of dominating probabilities has vertices associated with permutations of the list of element of Θ. This results suggests that the set of dominating k-additive belief functions should have a similar form, with each vertex associated with a permutation of the list of focal elements of size *smaller than or equal to k*. We first need to move from the probability simplex of all probability measure on Θ to the "belief space" of all belief measures on 2^Θ.

As shown in [46–48], it is possible to interpret belief functions as points of a Cartesian space, and study the interplay of Dempster-Shafer theory objects from a geometrical point of view.

Given a frame Θ, each belief function $b : 2^\Theta \rightarrow [0,1]$ is completely specified by its $N \doteq 2^{|\Theta|} - 2$ belief values $\{b(A), \text{ such that } \emptyset \subsetneq A \subsetneq \Theta\}$, (as $b(\emptyset) = 0$, $b(\Theta) = 1 \; \forall b$) and can therefore be represented as a vector of $\mathbb{R}^N$:

$$\vec{b} = [b(A), \emptyset \subsetneq A \subsetneq \Theta]' \tag{27}$$

If we denote by b_A the *categorical* belief function [49] assigning all the mass to a single subset $A \subseteq \Theta$, $m_{b_A}(A) = 1$, $m_{b_A}(B) = 0 \; \forall B \subseteq \Theta$, $B \neq A$, we can prove that [46,48] the set of points of $\mathbb{R}^N$ which correspond to a belief function, called the *belief space* and noted $\mathcal{B}$, coincides with the convex closure [2] Cl of all the vectors representing categorical belief functions: $\mathcal{B} = Cl(\vec{b}_A, \emptyset \subsetneq A \subseteq \Theta)$. The belief space $\mathcal{B}$ is a simplex [48], and each vector $\vec{b} \in \mathcal{B}$ representing a belief function b can be written as a convex sum as:

$$\vec{b} = \sum_{\emptyset \subsetneq B \subseteq \Theta} m_b(B) \vec{b}_B. \tag{28}$$

Let us then denote by $\mathcal{B}_k$ the region of the belief space $\mathcal{B}$ associated with k-additive belief functions. Clearly, the set $\mathcal{B}_1$ of all Bayesian belief functions on Θ is the simplex determined by all basis belief functions associated with singletons: $\mathcal{B}_1 = Cl(b_x, x \in \Theta)$.

In [31], the authors consider the dominance properties of k-additive belief functions for any type of capacities [32], define $\mathfrak{B}^k[b]$ as the polytope of k-additive belief functions dominating another belief function, and provide some results to characterize it.

Here we proceed to a full description of the polytope $\mathfrak{B}^k[b]$ in terms of its vertices.

---

[2] $Cl(\vec{b}_1, ..., \vec{b}_k) = \{\vec{b} \in \mathcal{B} : \vec{b} = \alpha_1 \vec{b}_1 + \cdots + \alpha_k \vec{b}_k, \sum_i \alpha_i = 1, \; \alpha_i \geq 0 \; \forall i\}$.

6.3 Vertices of the polytope and permutations of focal elements

Theorem 4 *Given a belief function $b : 2^{\Theta} \rightarrow [0,1]$, with mass function m, the region $\mathfrak{B}^k[b]$ of all the k-additive belief functions on Ω which dominate b according to order relation (7) is the polytope:*

$$\mathfrak{B}^k[b] = Cl(b^{\rho}[b] \; \forall \rho),$$

where ρ is any permutation $\{A_{\rho(1)}, ..., A_{\rho(|\mathcal{P}^k(\Omega)|)}\}$ of the focal elements of Ω of size at most k ($\mathcal{P}^k(\Omega)$), and the vertex $b^{\rho}[b]$ is the k-additive belief function with the following mass function:

$$m^{\rho}[b](A_{\rho(i)}) = \sum_{B \supseteq A_{\rho(i)};\; B \not\supset A_{\rho(j)} \; \forall j<i} m(A). \qquad (29)$$

Moreover, each vertex $b^{\rho}[b]$ of $\mathfrak{B}^k[b]$ is associated with the same number of permutations of $\mathcal{P}^k(\Omega)$.

Proof. □

7 Conclusions

Belief functions possess a strong credal semantics in terms of convex sets of probability distributions or consistent probabilities, for whose values on all events belief values provide lower bounds. These probabilities can also be seen as more committed or “inner” Bayesian approximations of the original b.f.
In this paper we proved a more compact form of the polytope of consistent probabilities, as the latter has $n!$ (candidate) vertices each corresponding to a different permutation of the elements of the domain. This unveils an interesting link with the vertices of the polytopes of outer consonant approximations also generated by permutations of singletons both in terms of their analytical expression and their convex geometry.

References

[1] D. Mercier, T. Denoeux, M. Masson, Refined sensor tuning in the belief function framework using contextual discounting, in: IPMU, 2006.

[2] S. Demotier, W. Schon, T. Denoeux, Risk assessment based on weak information using belief functions: a case study in water treatment, IEEE Transactions on Systems, Man and Cybernetics, Part C 36(3) (May 2006) 382– 396.

[3] G. Shafer, A Mathematical Theory of Evidence, Princeton University Press, 1976.

[4] D. Dubois, H. Prade, Possibility theory, Plenum Press, New York, 1988.

[5] T. Denoeux, A new justification of the unnormalized Dempsters rule of combination from the Least Commitment Principle, in: Proceedings of FLAIRS'08, Special Track on Uncertaint Reasoning, 2008.

[6] R. R. Yager, The entailment principle Dempster-Shafer granules, International Journal of Intelligent Systems 1 (1986) 247–262.

[7] D. Dubois, H. Prade, A set-theoretic view of belief functions: logical operations and approximations by fuzzy sets, International Journal of General Systems 12 (1986) 193–226.

[8] I. Levi, The enterprise of knowledge, MIT Press, 1980.

[9] G. Shafer, Constructive probability, Synthese 48 (1981) 309–370.

[10] A. Dempster, Lindley's paradox: Comment, Journal of the American Statistical Association 77:378 (June 1982) 339–341.

[11] P. Smets, The nature of the unnormalized beliefs encountered in the transferable belief model, in: Proceedings of the 8th Annual Conference on Uncertainty in Artificial Intelligence (UAI-92), Morgan Kaufmann, San Mateo, CA, 1992, pp. 292–29.

[12] D. Dubois, M. Grabisch, H. Prade, P. Smets, Using the transferable belief model and a qualitative possibility theory approach on an illustrative example: the assessment of the value of a candidate, Intern. J. Intell. Systems.

[13] P. Snow, The vulnerability of the transferable belief model to dutch books, Artificial Intelligence 105 (1998) 345–354.

[14] P. Smets, Belief functions versus probability functions, in: S. L. Bouchon B., Y. R. (Eds.), Uncertainty and Intelligent Systems, Springer Verlag, Berlin, 1988, pp. 17–24.

[15] P. Smets, R. Kennes, The transferable belief model, Artificial Intelligence 66 (1994) 191–234.

[16] P. Smets, Decision making in the TBM: the necessity of the pignistic transformation, International Journal of Approximate Reasoning 38 (2) (February 2005) 133–147.

[17] A. Chateauneuf, J. Y. Jaffray, Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion, Mathematical Social Sciences 17 (1989) 263–283.

[18] T. Melkonyan, R. Chambers, Degree of imprecision: Geometric and algebraic approaches, forthcoming in the International Journal of Approximate Reasoning.

[19] F. G. Cozman, Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions, Journal of Computational and Graphical Statistics 8(4) (1999) 824–838.

[20] Berger, Robust bayesian analysis: Sensitivity to the prior, Journal of Statistical Planning and Inference 25 (1990) 303–328.

[21] T. Herron, T. Seidenfeld, L. Wasserman, Divisive conditioning: further results on dilation, Philosophy of Science 64 (1997) 411–444.

[22] T. Seidenfeld, L. Wasserman, Dilation for convex sets of probabilities, Annals of Statistics 21 (1993) 1139–1154.

[23] V. Ha, P. Haddawy, Geometric foundations for interval-based probabilities, in: A. G. Cohn, L. Schubert, S. C. Shapiro (Eds.), KR'98: Principles of Knowledge Representation and Reasoning, Morgan Kaufmann, San Francisco, California, 1998, pp. 582–593.
URL `citeseer.ist.psu.edu/ha98geometric.html`

[24] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.

[25] E. Miranda, I. Couso, P. Gil, Extreme points of credal sets generated by 2-alternating capacities, International Journal of Approximate Reasoning 33 (2003) 95–115.

[26] A. Dempster, Upper and lower probabilities induced by a multivariate mapping, Annals of Mathematical Statistics 38 (1967) 325–339.

[27] L. Shapley, Cores of convex games, Int. J. Game Theory 1 (1971) 11–26.

[28] A. Chateauneuf, J. Jaffray, Some characterization of lower probabilities and other monotone capacities through the use of mbius inversion, Math. Soc. Sci. 17 (1989) 263–283.

[29] A. Wallner, Maximal number of vertices of polytopes defined by f-probabilities, in: F. G. Cozman, R. Nau, T. Seidenfeld (Eds.), ISIPTA 2005 – Proceedings of the Fourth International Symposium on Imprecise Probabilities and Their Applications, SIPTA, 2005, pp. 126–139.

[30] P. Miranda, M. Grabisch, P. Gil, On some results of the set of dominating k-additive belief functions, in: IPMU, 2004, pp. 625–632.

[31] P. Miranda, M. Grabisch, P. Gil, Dominance of capacities by k-additive belief functions, European Journal of Operational Research 175 (2006) 912–930.

[32] P. Walley, Towards a unified theory of imprecise probability, International Journal of Approximate Reasoning 24 (2000) 125–148.

[33] P. Baroni, Extending consonant approximations to capacities, in: Proceedings of IPMU, 2004, pp. 1127–1134.

[34] D. Dubois, H. Prade, Consonant approximations of belief functions, International Journal of Approximate Reasoning 4 (1990) 419–449.

[35] A. Dempster, A generalization of Bayesian inference, in: Classic Works of the Dempster-Shafer Theory of Belief Functions, 2008, pp. 73–104.

[36] P. Smets, Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem, International Journal of Approximate Reasoning 9 (1993) 1–35.

[37] T. Denoeux, Conjunctive and disjunctive combination of belief functions induced by non distinct bodies of evidence, Artificial Intelligence (to appear).

[38] F. Cuzzolin, A geometric approach to the theory of evidence, IEEE Trans. Systems, Man, and Cybernetics C (in press) 38 (3).

[39] A. Antonucci, F. Cuzzolin, Credal sets approximation by lower probabilities: Application to credal networks, in: Proc. of IPMU 2010, 2010.

[40] F. G. Cozman, Credal networks, Artificial Intelligence 120 (2000) 199–233.

[41] P. Walley, Statistical Reasoning with Imprecise Probabilities, Chapman and Hall, London, 1991.

[42] D. Dubois, H. Prade, P. Smets, New semantics for quantitative possibility theory., in: ISIPTA, 2001, pp. 152–161.

[43] C. Joslyn, Towards an empirical semantics of possibility through maximum uncertainty, in: R. Lowen, M. Roubens (Eds.), Proc. IFSA 1991, Vol. A, 1991, pp. 86–89.

[44] F. Cuzzolin, The geometry of consonant belief functions: simplicial complexes of necessity measures, Fuzzy Sets and Systems 161 (10) (2010) 1459–1479.

[45] M. Grabisch, K-order additive discrete fuzzy measures and their representation, Fuzzy sets and systems 92 (1997) 167–189.

[46] F. Cuzzolin, R. Frezza, Geometric analysis of belief space and conditional subspaces, in: Proceedings of the 2^{nd} International Symposium on Imprecise Probabilities and their Applications (ISIPTA2001), Cornell University, Ithaca, NY, 26-29 June 2001.

[47] F. Cuzzolin, Visions of a generalized probability theory, PhD dissertation, Università di Padova, Dipartimento di Elettronica e Informatica (19 February 2001).

[48] F. Cuzzolin, A geometric approach to the theory of evidence, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews 38 (4) (2008) 522–534.

[49] P. Smets, R. Kennes, The transferable belief model, Artificial intelligence 66 (2) (1994) 191–234.