Geometry of Upper Probabilities*

F. CUZZOLIN
Università di Padova, Italy
Politecnico di Milano, Italy

Abstract

In this paper we adopt the geometric approach to the theory of evidence to study the geometric counterparts of the plausibility functions, or upper probabilities. The computation of the coordinate change between the two natural reference frames in the belief space allows us to introduce the dual notion of basic plausibility assignment and understand its relation with the classical basic probability assignment. The convex shape of the plausibility space Π is recovered in analogy to what was done for the belief space, and the pointwise geometric relation between a belief function and the corresponding plausibility vector is discussed. The orthogonal projection of an arbitrary belief function s onto the probabilistic subspace is computed and compared with other significant entities, such as the relative plausibility and mean probability vectors.

Keywords

theory of evidence, belief space, basic plausibility assignment, plausibility space, orthogonal projection

1 Introduction

Uncertainty measures are assuming a mayor role in fields like artificial intelligence and computer vision, where problems requiring formalized reasoning are common. However, during the last decades a number of different descriptions of uncertain state of knowledge have been proposed, as alternatives or extensions of the classical probability theory. The theory of evidence is one of the most popular formalisms, thanks perhaps to its nature of quite natural extension of the classical Bayesian methodology.

In a series of recent works ([7], [6]) we have proposed a geometric interpretation of the theory of evidence based on the notion of *belief space*, the set of all

*This work has been supported by the Autonomous Navigation and Computer Vision Lab, Department of Information Engineering, led by professor R. Frezza.

the b.f.s defined on a fixed domain. It is well known that upper and lower probabilities, belief functions, possibility measures, fuzzy sets can be all thought of as *fuzzy measures*. Hence, it would be highly desirable to find a common environment where to discuss and compare all these uncertainty descriptions in an unified fashion.

In this perspective, this paper proposes a geometric picture of the connections between upper and lower probabilities in the belief space framework. After recalling the basic notions of the theory of evidence, we will briefly introduce the geometric approach to the ToE. After computing the change of coordinates between the orthogonal and oblique reference frames in the belief space, the notion of basic plausibility assignment will be defined and its analytic relation with the basic probability assignment unveiled (Section 3). This will allow us to describe the space of all the plausibility vectors as a simplex, called *plausibility space*, and give a natural interpretation of its vertices in terms of degrees of belief.

Next (Section 4) we will try and understand the pointwise geometry of upper probabilities by noticing that the line connecting a belief function s and the corresponding plausibility function P_s^* is *orthogonal* to the Bayesian subspace $\mathcal{P}$. This will allow us to compute the *orthogonal projection* $s_{\perp\mathcal{P}}$ of s onto $\mathcal{P}$ and prove that it is a probability distribution. We will also find the position of the mean probability vector $\frac{s+P_s^*}{2}$ and the condition under which P_s^* is the reflection of s through the probabilistic subspace.

Finally, we will express the credal set of the probabilities consistent with s as a simplex, noticing that its center of mass is the geometric counterpart of the so called *pignistic* transformation, and discuss the geometry of these points in the perspective of the probabilistic approximation problem. To improve the readability of the paper the proofs of the major results have been moved to an appendix.

1.1 Previous work

The geometric approach to the theory of evidence and generalized probabilities is due to the author, even if close references can be the works of Ha and Haddawy [9] and Wang *et al.* [17]. Anyway, some interesting papers have been recently published on the geometry of lower probabilities and plausibilities of singletons. P. Black, in particular, has dedicated its doctoral thesis to the study of belief functions [2]. An abstract of his results on the geometry of belief functions and other monotone capacities can be found in [3], where he uses shapes of geometric loci to give a direct visualization of the distinct classes of monotone capacities. In particular a number of results about lengths of edges of convex sets representing monotone capacities are given, together with their *size* meant as the sum of those lengths.

A number of papers, on the other side, have been published on the approximation of belief functions (see [1] for a review), mainly in order to find efficient implementations of the rule of combination aiming to reduce the number of focal elements (see for instance the works of Tessem [16] and Lowrance *et al.* [11]).

2 Geometric approach to the Theory of Evidence

The *theory of evidence* [13] has been introduced in the late Seventies by Glenn Shafer as a way of representing epistemic knowledge, starting from a sequence of seminal works of Arthur Dempster [8]. In this formalism the best representation of chance is a *belief function* (b.f.) rather than a Bayesian mass distribution. Following Shafer [13] let us call the finite set of possible outcomes for a decision problem *frame of discernment* or simply *frame*. In the following we will denote by A^c the complement of an arbitrary set A, by $A \setminus B \doteq A \cap B^c$ the difference of two sets A and B, and by $|A|$ the cardinality (number of elements) of A.

A *basic probability assignment* (b.p.a.) over a frame Θ is a function $m : 2^\Theta \rightarrow [0,1]$ on its power set $2^\Theta = \{A \subset \Theta\}$ such that

$$m(\emptyset) = 0, \; \sum_{A \subset \Theta} m(A) = 1, \;\; m(A) \geq 0 \;\; \forall A \subset \Theta.$$

The subsets of Θ associated with non-zero values of m are called *focal elements* and their union $\mathcal{C}$ *core*.
The *belief function* $s : 2^\Theta \rightarrow [0,1]$ associated with a basic probability assignment m is defined as $s(A) = \sum_{B \subset A} m(B)$, while m can be uniquely recovered from s by means of the *Moebius formula*

$$m(A) = \sum_{B \subset A} (-1)^{|A \setminus B|} s(B). \tag{1}$$

In particular, a *Bayesian* belief function s is a belief function such that $m_s(A) = 0$ for all A s.t. $|A| > 1$. Hence, finite probabilities are nothing more than special b.f.s.

Belief functions representing distinct bodies of evidence can be combined by means of the *Dempster's rule of combination* [8]. The *orthogonal sum* $s_1 \oplus s_2$ of two belief functions is a new belief function whose focal elements are all the possible intersections between the combining focal elements and whose b.p.a. is given by

$$m(C) = \frac{\sum_{i,j: A_i \cap B_j = C} m_1(A_i) m_2(B_j)}{1 - \sum_{i,j: A_i \cap B_j = \emptyset} m_1(A_i) m_2(B_j)}. \tag{2}$$

where $\{A_i\}$ and $\{B_j\}$ are the focal elements of s_1, s_2 respectively.
When all the intersections between focal elements of the two functions are empty, the denominator of Equation (2) goes to zero and we say that s_1 and s_2 are *not combinable*.

A *dual* representation of the evidence encoded by a belief function s is called *upper probability*[1], and expresses the amount of evidence *not against* a proposi-

[1] The name comes from the fact that belief values and upper probability values are respectively lower and upper bounds for the probabilities of the events.

tion A

$$P^*(A) \doteq 1 - s(A^c) = 1 - \sum_{B\subset A^c} m(B) = \sum_{B\cap A\neq\emptyset} m(B) \geq s(A). \tag{3}$$

Now, consider a frame of discernment Θ and introduce in the Euclidean space $\mathbb{R}^{|2^\Theta|-1}$ an orthonormal reference frame $\{X_A\}_{A\subset\Theta, A\neq\emptyset}$ such that each coordinate function x_A measures the belief value associated with the i-th subset of Θ.

Definition 1 *The* belief space *associated with Θ is the set of points $\mathcal{S}_\Theta$ of $\mathbb{R}^{|2^\Theta|-1}$ corresponding to a belief function.*

We usually assume the domain Θ fixed, and denote the belief space by $\mathcal{S}$. Let us call A-th *basis belief function*

$$P_A \doteq s \in \mathcal{S} \; s.t. \; m_s(A) = 1, \; m_s(B) = 0 \; B \neq A$$

the unique belief function assigning all the mass to a single subset A of Θ. It can be proved that (see [7], [6]), calling $\mathcal{E}_s$ the list of focal elements of s,

Theorem 1 *The set of all the belief functions with focal elements in a given collection X is closed and convex in $\mathcal{S}$: $\{s : \mathcal{E}_s \subset X\} = Cl(\{P_A : A \in X\})$.*

The shape of $\mathcal{S}$ follows immediately from Theorem 1.

Corollary 1 *The belief space $\mathcal{S}$ coincides with the convex closure of all the basis belief functions, $\mathcal{S} = Cl(P_A, \, A \subset \Theta, \, A \neq \emptyset)$.*

Moreover, any belief function $s \in \mathcal{S}$ can be written as a convex sum as follows:

$$s = \sum_{A\subset\Theta, \, A\neq\emptyset} m_s(A) \cdot P_A. \tag{4}$$

Clearly, since a probability is a belief function assigning non zero masses to singletons only, Theorem 1 implies that the set $\mathcal{P}$ of all the Bayesian belief functions is a subset of the border of $\mathcal{S}$, precisely $\mathcal{P} = Cl(P_{\{\theta_i\}}, \; i = 1, ..., |\Theta|)$.

3 Geometry of Plausibility Functions

Analogously to what done for the vectors of $\mathbb{R}^N$ ($N \doteq |2^\Theta| - 1$) representing belief functions, we would like to understand the geometric properties of the plausibility vectors $[P^*_s(A), A \subset \Theta]'$. A plausibility vector can indeed be expressed as

$$P^*_s = \sum_{A\subset\Theta} P^*_s(A) \cdot X_A \tag{5}$$

where $\{X_A, A \subset \Theta\}$ is the orthogonal reference frame of the belief space.
The basis belief functions P_A form a set of independent vectors in $\mathbb{R}^N$, so that the

collections $\{X_A\}$ and $\{P_A\}$ form two distinct coordinate frames in the belief space. To understand the place a plausibility vector takes in the belief reference frame $\{P_A\}$ we then need to compute the coordinate change between these frames. We first notice that basis b.f.s can be expressed as $P_A = \sum_{E \supset A} X_E$.

Proposition 1 *The coordinate change between the two coordinate frames* $\{X_A\}$ *and* $\{P_A\}$ *is given by*

$$X_A = \sum_{B \supset A} P_B \cdot (-1)^{|B \setminus A|}. \tag{6}$$

3.1 Basic Plausibility Assignment

Let us now replace expression (6) in Equation (5), obtaining for P_s^*[2]

$$\sum_{A \subset \Theta} P_s^*(A) \cdot X_A = \sum_{A \subset \Theta} P_s^*(A) \cdot \sum_{B \supset A} P_B \cdot (-1)^{|B \setminus A|} = \sum_{B \subset \Theta} P_B \cdot \sum_{A \subset B} (-1)^{|B-A|} P_s^*(A)$$

and after introducing the quantity

$$\mu(A) \doteq \sum_{B \subset A} (-1)^{|A-B|} P_s^*(B) \tag{7}$$

we can write

$$P_s^* = \sum_{A \subset \Theta} \mu(A) \cdot P_A. \tag{8}$$

We call the function $\mu : 2^\Theta \to \mathbb{R}$ defined by expression (7) *basic plausibility assignment*. It is easy to recognize the Moebius equation for plausibilities, which implies $P_s^*(A) = \sum_{B \subset A} \mu(B)$. A few calculations allow us to understand the relation between basic probabilities and plausibilities.

Theorem 2

$$\mu(A) = \begin{cases} (-1)^{|A|+1} \sum_{E \supset A} m(E) & A \neq \emptyset \\ 0 & A = \emptyset. \end{cases} \tag{9}$$

It is easy to see that basic plausibility assignments *meet the normalization constraint*. In fact

$$\sum_{A \subset \Theta} \mu(A) = -\sum_{A \subset \Theta, A \neq \emptyset} (-1)^{|A|} \sum_{E \supset A} m(E) = -\sum_{E \subset \Theta} m(E) \cdot \sum_{A \subset E, A \neq \emptyset} (-1)^{|A|} = 1$$

since $-\sum_{A \subset E, A \neq \emptyset} (-1)^{|A|} = -(0 - (-1)^0) = 1$ for the expression of Newton's binomial $\sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n$, where in this case $k = |A|$, $p = -1$, $q = 1$. However, $\mu(A)$ is not always positive, so we can just infer that any plausibility vector lies on the affine subspace generated by the basis belief functions $\{P_A\}$.

[2]Note that $P_s^*(\emptyset) = 0$ so the expression is correct even if $X_\emptyset$ does not exist.

3.2 Plausibility Space

Analogously to what done for belief functions, let us call *plausibility space* the region Π of $\mathbb{R}^N$ whose points correspond to admissible plausibility functions. It is not difficult to prove that

Theorem 3 Π *is a simplex*

$$\Pi = Cl(\Pi_A, A \subset \Theta, A \neq \emptyset), \quad \Pi_A = -\sum_{B \subset A} (-1)^{|B|} P_B. \tag{10}$$

Proof. We just need to re-assemble expression (8) as a convex combination of points, getting (through Equation (9))

$$\begin{aligned} P_s^* &= \sum_{A \subset \Theta} \mu(A) \cdot P_A = \sum_{A \subset \Theta, A \neq \emptyset} (-1)^{|A|+1} \cdot \sum_{E \supset A} m(E) \cdot P_A = \\ &= \sum_{A \subset \Theta, A \neq \emptyset} \sum_{E \supset A} (-1)^{|A|+1} m(E) \cdot P_A = \sum_{E \subset \Theta, E \neq \emptyset} m(E) \cdot \sum_{A \subset E, A \neq \emptyset} (-1)^{|A|+1} P_A \end{aligned}$$

$= \sum_{E \neq \emptyset} m(E) \Pi_E$, that is a convex combination since basic probability assignments have unitary sum. □

It is easy to notice that $\Pi_{\{\theta\}} = -(-1)^{|\{\theta\}|} \cdot P_{\{\theta\}} = P_{\{\theta\}} \forall \theta \in \Theta$, so that $\mathcal{P} \subset \mathcal{S} \cap \Pi$. The inverse relation between basis belief functions and basis plausibilities has the same form of Equation (10):

Theorem 4

$$P_A = -\sum_{B \subset A} (-1)^{|B|} \cdot \Pi_B. \tag{11}$$

Proof. The proof follows the sketch of Proposition 1. Replacing expression (11) in Equation (10) yields for Π_A

$$-\sum_{B \subset A} (-1)^{|B|} P_B = \sum_{B \subset A} (-1)^{|B|} \cdot \sum_{E \subset B} (-1)^{|E|} \Pi_E = \sum_{E \subset A} (-1)^{|E|} \Pi_E \cdot \sum_{E \subset B \subset A} (-1)^{|B|}$$

but then, analogously to what previously done (see the Appendix),

$$\sum_{E \subset B \subset A} (-1)^{|B|} = \begin{cases} (-1)^{|A|} & E = A \\ 0 & E \neq A \end{cases}$$

and the thesis easily follows. □

The vertices of the plausibility space have a natural interpretation.

Theorem 5 *The vertex Π_A of the plausibility space is the plausibility vector associated with the basis belief function P_A, $\Pi_A = P^*_{P_A}$.*

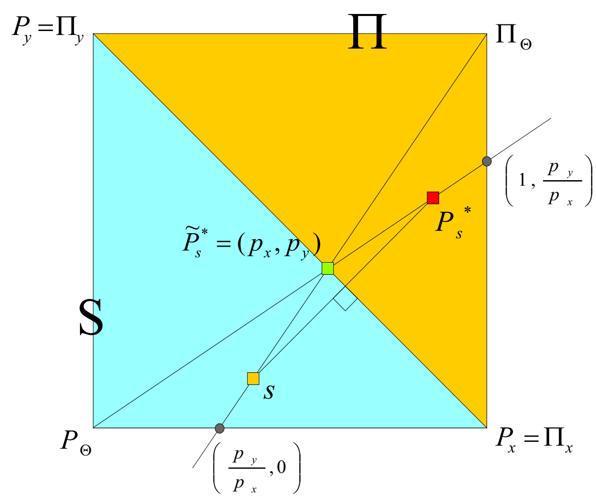

Figure 1: Geometric relations between upper and lower probabilities in the belief space for a binary frame $\Theta = \{x,y\}$. The belief space $\mathcal{S}$ and the plausibility space Π are both simplices with vertices $\{P_\Theta = (0,0), P_x = (1,0), P_y = (0,1)\}$ and $\{\Pi_\Theta = (1,1), \Pi_x = P_x, \Pi_y = P_y\}$ respectively. In the picture a belief function s and the corresponding plausibility function P_s^* are indicated, showing that they are in symmetric positions with respect to the common subspace $\mathcal{P}$. The location of the relative plausibility of singletons $\tilde{P}_s^*$ is also shown, as intersection of the probabilistic subspace with the line joining P_s^* and $P_\Theta = (0,0)$. A dual line joining s and Π_Θ also appears.

Figure 1 shows the relation between belief and plausibility space for a the binary frame $\Theta = \{x,y\}$. Without reporting the calculations, we may notice another few interesting facts. The two simplices are perfectly symmetric with respect to the probabilistic subspace. Furthermore, upper and lower probability vectors determine a line that is orthogonal to $\mathcal{P}$, and they also lie on symmetric positions with respect to the Bayesian region. Notice that the relative plausibility vector $\tilde{P}_s^*$ (normalized version of P_s^*) does not coincide at all with the orthogonal projection of s (or P_s^*) onto $\mathcal{P}$. In the following we will try and understand what of those features retain their validity in the general case.

4 Upper and lower probability vectors

It is in fact natural to wonder what is the pointwise relation between vectors representing upper and lower probability functions generated by the same evidence.

Luckily enough, orthogonality turns out to be an actual property of those uncertainty descriptions.

4.1 Orthogonal projection

Let us first denote with P_x the basis belief function for $A = \{x\}$. Being $\mathcal{P} = Cl(P_x, x \in \Theta)$ an affine subspace, it can be written as the translated version of a vector space as $\mathcal{P} = P_x + span(P_y - P_x, \forall y \in \Theta, y \neq x)$, where the $n-1$ vectors $P_y - P_x$ form a basis of this vector space. They show a peculiar symmetry

$$P_y - P_x(A) = \begin{cases} 1 & A \supset \{y\}, A \not\supset \{x\} \\ 0 & A \supset \{x\}, \{y\} \; or \; A \not\supset \{x\}, \{y\} \\ -1 & A \not\supset \{y\}, A \supset \{x\}. \end{cases}$$

that can be usefully exploited for our goals. In particular, we can appreciate that

$$(P_y - P_x)(A) = 1 \Rightarrow A \supset \{y\}, A \not\supset \{x\} \Rightarrow A^c \supset \{x\}, A^c \not\supset \{y\} \Rightarrow (P_y - P_x)(A^c) = -1$$

and vice-versa, while $(P_y - P_x)(A) = 0 \Rightarrow A \supset \{y\}, A \supset \{x\}$ or $A \not\supset \{y\}, A \not\supset \{x\}$ so that in the first case $A^c \not\supset \{x\}, \{y\}$, in the second one $A^c \supset \{x\}, \{y\}$ but in both situations $(P_y - P_x)(A^c) = 0$. Summarizing we can write

$$(P_y - P_x)(A^c) = -(P_y - P_x)(A) \quad \forall A \subset \Theta$$

which directly implies that

Theorem 6 *The line connecting P_s^* and s is orthogonal to the probabilistic subspace, i.e.*

$$s - P_s^* \perp \mathcal{P}.$$

It is then clear that the orthogonal projection of s onto $\mathcal{P}$ is simply the intersection of this line with the probabilistic subspace,

$$s_{\perp\mathcal{P}} = \vec{s}P_s^* \cap \mathcal{P}.$$

We just have to find the value of α such that $s + \alpha(P_s^* - s) \in \mathcal{P}$.

Theorem 7 *The coordinates of the orthogonal projection of s onto $\mathcal{P}$ with respect to the basis $\{P_A\}$ can be expressed in terms of the basic probability assignment m of s as follows:*

$$m_{s_{\perp\mathcal{P}}}(\{x\}) = m(\{x\}) + \sum_{A \supseteq \{x\}} m(A) \cdot \frac{\sum_{|A|>1} m(A)}{\sum_{|A|>1} m(A)|A|}. \tag{12}$$

Equation (12) ensures that $m_{s_{\perp\mathcal{P}}}(\{x\})$ is always positive for each $x \in \Theta$, so that

Corollary 2 *The orthogonal projection $s_{\perp\mathcal{P}}$ of any arbitrary belief function s onto the probabilistic subspace $\mathcal{P}$ is a Bayesian belief function.*

This fact is not just a trivial consequence of its definition, since the probability simplex is a small region of $span(\mathcal{P})$ in general. A symmetric version of the formula can be obtained after realizing that $\frac{\sum_{|A|=1} m(A)}{\sum_{|A|=1} m(A)|A|} = 1$, so that we can write

$$m_{s_{\perp\mathcal{P}}}(\{x\}) = s(\{x\}) \cdot \frac{\sum_{|A|=1} m(A)}{\sum_{|A|=1} m(A)|A|} + [P_s^* - s](\{x\}) \cdot \frac{\sum_{|A|>1} m(A)}{\sum_{|A|>1} m(A)|A|}. \quad (13)$$

It is natural to wonder whether the upper probability vector is actually the reflection of the lower probability vector through the probabilistic subspace as in the binary case, i.e. if $s_{\perp\mathcal{P}} = \frac{s + P_s^*}{2}$. In [5] we will show that

Proposition 2 *Orthogonal projection and mean probability coincide iff*

$$\sum_{|A|>1} m(A)|A| = 2 \sum_{|A|>1} m(A).$$

This apparently arid result is strictly related to the duality isuue concerning the geometric counterparts of upper and lower probabilities. Is this duality associated with some kind of symmetry through the probabilistic subspace? Further analysis [5] seem to hint that the situation is a bit more complex.

4.2 Simplex of Consistent Probabilities

It is well known, on the other side, that belief functions can be formally interpreted in terms of classes of unknown probabilities. Given the nature of basic probability assignments, it is natural to conjecture that the set of probabilities $P(s)$ consistent with a given belief function s has also the shape of a simplex. Is there any relation between the orthogonal projection of s onto $\mathcal{P}$ and this simplex?

Following Shafer [13] we can think of $m(A)$ as a probability free to move inside A. If we assign the mass of each focal element A_i to one of its elements a_i, intuitively we should get an extremum of the region of consistent probabilities. More formally, to each focal element A corresponds a mass $m(A)$ distributed among its elements, $m(A) \cdot Cl(P_a,\, a \in A)$, so that $P(s)$ can be expressed as

$$P(s) = \sum_{A \subset \Theta} m(A) \cdot Cl(P_a,\, a \in A).$$

Then, given an arbitrary belief function s with focal elements $A_1, ..., A_m$, we can define for each choice of m representatives $\{a_1, ..., a_m\}$, $a_i \in A_i\ \forall i$,

$$P_{a_1 \ldots a_m} \doteq \sum_{i=1}^{m} m(A_i) \cdot P_{a_i}. \quad (14)$$

It can be proved that [5] (as suggested by our intuition)

Proposition 3

$$P(s) = Cl(P_{a_1 \dots a_m}, \{a_1, \dots, a_m\} \in A_1 \times \dots \times A_m).$$

Accordingly, the center of mass $\bar{P}(s)$ of $P(s)$ gets the form

$$\frac{1}{\prod_i |A_i|} \cdot \sum_{\{a_1,\dots,a_m\} \in A_1 \times \dots \times A_m} P_{a_1 \dots a_m} = \frac{1}{\prod_i |A_i|} \cdot \sum_{\{a_1,\dots,a_m\} \in A_1 \times \dots \times A_m} \sum_{i=1}^{m} m(A_i) P_{a_i} =$$
$$\frac{1}{\prod_i |A_i|} \sum_{a \in C_s} P_a \sum_{A_j \supset \{a\}} m(A_j) \frac{\prod_i |A_i|}{|A_j|} = \sum_{a \in C_s} P_a \sum_{A_j \supset \{a\}} \frac{m(A_j)}{|A_j|} = \sum_{x \in \Theta} P_x \sum_{A \supset \{x\}} \frac{m(A)}{|A|} \quad (15)$$

since no focal elements include points outside the core. Equation (15) possesses several interesting interpretations.

4.2.1 Center of mass and pignistic transformation

In his popular *transferable belief model* [15] Philippe Smets has proposed an approach to the theory of evidence in which beliefs are represented at credal level (as convex sets of probabilities or belief functions), while decisions are made by resorting to a probabilistic approximation of belief function called *pignistic transformation* (see for instance [4]). Smets justifies his transformation by means of a so-called "rationality" requirement, which mathematically translates into a linearity constraint (see Theorem 3 of [14]).

It is pretty surprising to see that the pignistic transformation $Pign[s]$ of a belief function s is exactly expressed by Equation (15)

$$Pign[s](x) = \sum_{A \supset \{x\}} \frac{m(A)}{|A|},$$

making clear that the geometric counterpart of the pignistic transformation coincides with the center of mass of the simplex $P(s)$ of consistent probabilities. The full implications of this fact are still unclear, and deserve further investigations.

4.2.2 Consistency and Epsilon Contamination

The geometric analysis of the convex region of the consistent probabilities can be also related to a popular technique in robust statistics, the Epsilon Contamination Model. For a fixed $0 < \varepsilon < 1$ and a probability distribution P^*, the associated ε-contamination model is a convex class of distributions of the form $\{(1-\varepsilon)P^* + \varepsilon Q\}$ where Q is arbitrary.
Teddy Seidenfeld has proved that (for discrete domains) any ε-contamination model is equivalent to a belief function, whose corresponding consistent probabilities form the largest convex set induced by the collection of coherent lower

probabilities the model specifies for the elements of the domain (see [12], Theorem 2.10). It is worth noticing that in this special case P^* has the meaning of barycenter of the convex set, providing then another interesting interpretation of Equation (15).

5 Comments

What we have learned about the pointwise geometry of upper and lower probabilities can then be eventually depicted as in Figure 2. Each belief function s is associated with a simplex of consistent probabilities (the shaded triangle) $P(s)$ in the probabilistic subspace $\mathcal{P}$ (the larger triangle), whose center of mass $\bar{P}(s)$ (representing the pignistic transformation of s) is in general different from the orthogonal projection of s onto $\mathcal{P}$. The line $\overline{sP_s^*}$ is orthogonal to $\mathcal{P}$ but s and P_s^* are not on symmetric positions in general.

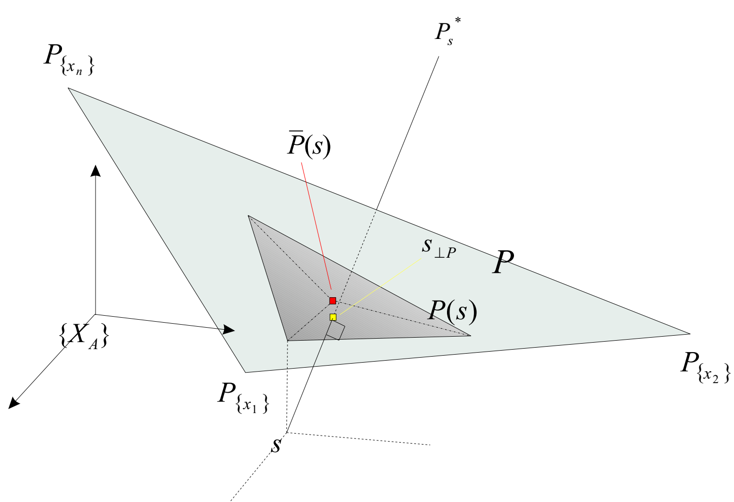

Figure 2: Geometric relation between upper and lower probability vectors.

The binary case turns out to be rather peculiar, since, recalling the definition of basic plausibility assignment (Section 3.1),

$$\begin{aligned}
\bar{P}(s) &= \textstyle\sum_{x\in\Theta_2} P_x \sum_{A\supset x} \frac{m(A)}{|A|} = P_x\cdot(m(x)+\frac{m(\Theta)}{2}) + P_y\cdot(m(y)+\frac{m(\Theta)}{2}),\\
\tfrac{s+P_s^*}{2} &= P_x\cdot\tfrac{m(x)+m(x)+m(\Theta)}{2} + P_y\cdot\tfrac{m(y)+m(y)+m(\Theta)}{2} + \\
&\quad + P_\Theta\cdot\tfrac{m(\Theta)-m(\Theta)}{2} = P_x\cdot(m(x)+\tfrac{m(\Theta)}{2}) + P_y\cdot(m(y)+\tfrac{m(\Theta)}{2}),\\
s_{\perp\mathcal{P}} &= P_x\cdot[m(x)+(1-m(y)-m(x))\cdot\tfrac{m(\Theta)}{2m(\Theta)}] + P_y\cdot[m(y)+\tfrac{1-m(x)-m(y)}{2}]\\
&= P_x\cdot(m(x)+\tfrac{m(\Theta)}{2}) + P_y\cdot(m(y)+\tfrac{m(\Theta)}{2})
\end{aligned}$$

and these three quantities coincide.

In our vision this knowledge could represent a step towards a more comprehensive understanding of the various uncertainty measures that can be introduced on finite domains: classical probabilities, upper and lower probabilities, belief functions, possibility measures, fuzzy sets. A number of papers have been recently published, for instance, on the connection between fuzzy measures and belief functions ([10] among the others). The belief space framework could provide a unifying environment where those connections may emerge more clearly and lead to a better comprehension of the field.

In this paper, in particular, we have seen how the dual concept of plausibility function or upper probability transfer into a dual convex geometry. The analogous of basis belief functions and probability assignments have been developed and their geometric interpretation exposed. We concentrated our efforts on understanding the pointwise relation between lower and upper probability vectors, proving their orthogonality with respect to the probabilistic subspace.

We also analyzed the comparative geometry of relative plausibility, orthogonal projection and center of mass of the set of consistent probabilities. This can be seen as a preliminary work in the perspective of a geometric solution to the probabilistic approximation problem. Coherently, we are also working on the geometry of finite fuzzy sets and possibility measures, to investigate more closely the idea of *duality* between probabilistic and possibilistic measures and discuss possible alternative consonant approximations of belief functions.

From a purely technical viewpoint, it is not clear yet what is the exact position in the belief space of a generic plausibility vector, and its geometric relation with other significant points like the relative plausibility of singletons $\tilde{P}_s^*$. In the next future [5] we will show how this quantity turns out to be the best Bayesian approximation of a belief function in the framework of Dempster's combination rule, and "perfectly" represents (in a very precise way) the original belief function in probabilistic subspace. It will be interesting to compare these findings with the results of a recent working paper Cobb and Shenoy [4], where they describe some properties of the relative plausibility of singletons and discuss its nature of probability function that is equivalent to the original belief function.

The study of consistent probabilities could play as well an important role in the search for an alternative to Dempster's rule of combination, for their description in terms of convex sets opens the way to the application of our commutativity results [6]. Understanding their behavior in an inference process could give us a hint of the properties a combination rule should possess to guarantee coherency in terms of the corresponding credal sets.

Appendix: Mathematical Proofs

Proof. (Proposition 1) If the thesis is true we have, by replacing X_A with expression (6),

$$P_A = \sum_{E \supset A} X_E = \sum_{E \supset A} \sum_{B \supset E} P_B \cdot (-1)^{|B-E|} = \sum_{B \supset A} P_B \cdot \sum_{B \supset E \supset A} (-1)^{|B-E|}.$$

Let us consider the factor $\sum_{A \subset E \subset B} (-1)^{|B-E|}$. When $A = B$ then $E = A = B$ and the coefficient becomes 1. On the other side, when $B \neq A$ we have

$$\sum_{A \subset E \subset B} (-1)^{|B-E|} = \sum_{F \subset B \setminus A} (-1)^{|B \setminus A \setminus F|} = 0$$

for Newton's binomial. Hence $P_A = P_A$. □

Proof. (Theorem 2) The definition (3) of upper probability yields

$$\begin{aligned} \mu(A) = \sum_{B \subset A} (-1)^{|A-B|} P_s^*(B) = \sum_{B \subset A} (-1)^{|A-B|} (1 - s(B^c)) = \\ = \sum_{B \subset A} (-1)^{|A-B|} - \sum_{B \subset A} (-1)^{|A-B|} s(B^c) \end{aligned} \tag{16}$$

where for Newton's binomial $\sum_{B \subset A} (-1)^{|A \setminus B|} = 0$ if $A \neq \emptyset$, $(-1)^{|A|}$ otherwise. If $B \subset A$ then $B^c \supset A^c$, so that the second addendum becomes

$$\begin{aligned} - \sum_{B \subset A, B \neq \emptyset} (-1)^{|A-B|} \sum_{E \subset B^c} m(E) = - \sum_{E \subset \Theta} m(E) \cdot \sum_{B: B \subset A, B^c \supset E} (-1)^{|A-B|} = \\ = - \sum_{E \subset \Theta} m(E) \cdot \sum_{B \subset A \cap E^c} (-1)^{|A-B|} \end{aligned} \tag{17}$$

for $B^c \supset E, B \subset A$ is equivalent to $B \subset E^c, B \subset A \equiv B \subset (A \cap E^c)$.
Let us now analyze the function of E

$$f(E) \doteq \sum_{B \subset A \cap E^c} (-1)^{|A-B|}.$$

If $A \cap E^c = \emptyset$ then $B = \emptyset$ and the sum is $(-1)^{|A|}$. If $A \cap E^c \neq \emptyset$, instead, we can write $F \doteq E^c \cap A$ and obtain (since $B \subset F \subset A$ and $|A-B| = |A-F| + |F-B|$)

$$f(E) = \sum_{B \subset F} (-1)^{|A-B|} = \sum_{B \subset F} (-1)^{|A-F|+|F-B|} = (-1)^{|A-F|} \cdot \sum_{B \subset F} (-1)^{|F-B|} = 0$$

given that $\sum_{B\subset F}(-1)^{|F-B|}=0$ for Newton's binomial again. Eventually

$$f(E)=\begin{cases}0 & E^c\cap A\neq\emptyset\\ (-1)^{|A|} & E^c\cap A=\emptyset.\end{cases}$$

We can then rewrite expression (17) as follows

$$\begin{array}{l}-\sum_{E\subset\Theta}m(E)f(E)=-\sum_{E:E^c\cap A\neq\emptyset}m(E)\cdot 0-\sum_{E:E^c\cap A=\emptyset}m(E)\cdot(-1)^{|A|}=\\ =(-1)^{|A|+1}\sum_{E:E^c\cap A=\emptyset}m(E)=(-1)^{|A|+1}\sum_{E\supset A}m(E)\end{array}$$

and replacing it in Equation (16) yields Equation (9), after distinguishing the two cases $A=\emptyset, A\neq\emptyset$. □

Proof. **(Theorem 5)** Expression (10) is equivalent to $\Pi_A(X)=-\sum_{B\subset A,B\neq\emptyset}(-1)^{|B|}P_B(X)\ \forall X\subset\Theta$. But since $P_B(X)=1$ if $X\supset B$ and 0 otherwise we have that

$$\Pi_A(X)=-\sum_{B\subset A,B\subset X,B\neq\emptyset}(-1)^{|B|}=-\sum_{B\subset A\cap X,B\neq\emptyset}(-1)^{|B|}.$$

Now, if $A\cap X=\emptyset$ there is no addenda in the above sum, that goes to zero. Otherwise, for Newton's binomial, we have $\Pi_A(X)=-\{[1+(-1)]^{|A\cap X|}-(-1)^0\}=1$. But then the definition of upper probability yields exactly

$$P^*_{P_A}(X)=\sum_{B\cap X\neq\emptyset}m_{P_A}(B)=\begin{cases}1 & A\cap X\neq\emptyset\\ 0 & A\cap X=\emptyset.\end{cases}$$

□

Proof. **(Theorem 6)** Clearly $P^*_s-s=\sum_{A\subset\Theta}X_A\cdot[P^*_s(A)-s(A)]$, where $[P^*_s-s](A^c)=P^*_s(A^c)-s(A^c)=1-s(A)-s(A^c)=1-s(A^c)-s(A)=P^*_s(A)-s(A)=[P^*_s-s](A)$. Hence,

$$\langle P^*_s-s,P_y-P_x\rangle=\sum_{A\subset\Theta}[P^*_s-s](A)\cdot[P_y-P_x](A)=$$

$$=\sum_{|A|\leq\lfloor|\Theta/2|\rfloor}[P^*_s-s](A)\cdot[(P_y-P_x)(A)-(P_y-P_x)(A^c)]=0$$

since $(P_y-P_x)(A)=-(P_y-P_x)(A^c)$. □

Proof. **(Theorem 7)** The desired condition implies that, for any subset $A\subset\Theta$, $s(A)+\alpha\cdot[P^*_s(A)-s(A)]=s(A)+\alpha\cdot[1-s(A^c)-s(A)]\in\mathcal{P}$. In particular, when $A=\{x\}$ is a singleton,

$$s(\{x\})+\alpha\cdot[1-s(\{x\}^c)-s(\{x\})]\in\mathcal{P}. \tag{18}$$

This point belongs to $\mathcal{P}$ iff the normalization criterion for singletons is met, i.e.

$$\sum_{x\in\Theta} s(\{x\}) + \alpha \cdot \sum_{x\in\Theta}(1-s(\{x\}^c)-s(\{x\})) = 1 \Rightarrow \alpha = \frac{1-\sum_{x\in\Theta}s(\{x\})}{\sum_{x\in\Theta}(1-s(\{x\}^c)-s(\{x\}))}$$

and after replacing this value of α into Equation (18) we get

$$s_{\perp\mathcal{P}}(\{x\}) = s(\{x\}) + \frac{1-\sum_{y\in\Theta}s(\{y\})}{\sum_{y\in\Theta}(1-s(\{y\}^c)-s(\{y\}))}\cdot(1-s(\{x\}^c)-s(\{x\})) =$$

$$= \frac{s(\{x\})\cdot[\sum_{y\in\Theta}(1-s(\{y\}^c)-s(\{y\}))-(1-\sum_{y\in\Theta}s(\{y\}))]}{\sum_{y\in\Theta}(1-s(\{y\}^c)-s(\{y\}))} +$$

$$+ \frac{(1-s(\{x\}^c))\cdot(1-\sum_{y\in\Theta}s(\{y\}))}{\sum_{y\in\Theta}(1-s(\{y\}^c)-s(\{y\}))} =$$

$$= \frac{s(\{x\})\cdot[\sum_{y\in\Theta}(1-s(\{y\}^c))-1]+(1-s(\{x\}^c))\cdot(1-\sum_{y\in\Theta}s(\{y\}))}{\sum_{y\in\Theta}(1-s(\{y\}^c)-s(\{y\}))}$$

that using the definition of plausibility function can be rewritten as

$$s_{\perp\mathcal{P}}(\{x\}) = \frac{s(\{x\})\cdot(\sum_{y\neq x}P_s^*(\{y\})-1)+P_s^*(\{x\})\cdot(1-\sum_{y\neq x}s(\{y\}))}{\sum_{y\in\Theta}[P_s^*(\{y\})-s(\{y\})]}. \quad (19)$$

Equation (19) determines the coordinate of the orthogonal projection of a belief function s onto $\mathcal{P}$. The expression for the basic probability assignment associated with this projection (Equation (12)) can be found after a few passages, extensively reported in [5]. □

References

[1] BAUER, M. Approximation algorithms and decision making in the Dempster-Shafer theory of evidence–an empirical study. *International Journal of Approximate Reasoning 17* (1997), 217–237.

[2] BLACK, P. *An examination of belief functions and other monotone capacities*. PhD dissertation, Department of Statistics, Carnegie Mellon University, 1996. Pgh. PA 15213.

[3] BLACK, P. Geometric structure of lower probabilities. In *Random Sets: Theory and Applications*, Goutsias, Malher, and Nguyen, Eds. Springer, 1997, pp. 361–383.

[4] COBB, B., AND SHENOY, P. On transforming belief function models to probability models. Tech. rep., University of Kansas, School of Business, Working Paper No. 293, February 2003.

[5] CUZZOLIN, F. Probabilistic approximations of belief functions. *in preparation*.

[6] CUZZOLIN, F. Geometrical structure of belief space and conditional subspaces. *submitted to the IEEE Transactions on Systems, Man and Cybernetics part C* (November 2002).

[7] CUZZOLIN, F., AND FREZZA, R. Geometric analysis of belief space and conditional subspaces. In *Proceedings of the* 2^{nd} *International Symposium on Imprecise Probabilities and their Applications (ISIPTA2001)* (26-29 June 2001).

[8] DEMPSTER, A. Upper and lower probabilities generated by a random closed interval. *Annals of Mathematical Statistics 39* (1968), 957–966.

[9] HA, V., AND HADDAWY, P. Theoretical foundations for abstraction-based probabilistic planning. In *Proc. of the* 12^{th} *Conference on Uncertainty in Artificial Intelligence* (August 1996), pp. 291–298.

[10] HEILPERN, S. Representation and application of fuzzy numbers. *Fuzzy Sets and Systems 91* (1997), 259–268.

[11] LOWRANCE, J. D., GARVEY, T. D., AND STRAT, T. M. A framework for evidential-reasoning systems. In *Proceedings of the National Conference on Artificial Intelligence* (1986), A. A. for Artificial Intelligence, Ed., pp. 896–903.

[12] SEIDENFELD, T. Some static and dynamic aspects of rubust Bayesian theory. In *Random Sets: Theory and Applications*, Goutsias, Malher, and Nguyen, Eds. Springer, 1997, pp. 385–406.

[13] SHAFER, G. *A Mathematical Theory of Evidence*. Princeton University Press, 1976.

[14] SMETS, P. Constructing the pignistic probability function in a context of uncertainty. In *Uncertainty in Artificial Intelligence, 5*, M. Henrion, R. Shachter, L. Kanal, and J. Lemmer, Eds. Elsevier Science Publishers, 1990, pp. 29–39.

[15] SMETS, P., AND KENNES, R. The transferable belief model. *Artificial Intelligence 66* (1994), 191–234.

[16] TESSEM, B. Approximations for efficient computation in the theory of evidence. *Artificial Intelligence 61:2* (1993), 315–329.

[17] WANG, C.-C., AND DON, H.-S. A geometrical approach to evidential reasoning. In *Proceedings of IEEE* (1991), pp. 1847–1852.

Fabio Cuzzolin is with the Department of Information Engineering, University of Padova, 35131 Padova Italy. E-mail: cuzzolin@dei.unipd.it