A Geometric Approach to the Theory of Evidence

Fabio Cuzzolin, *Member, IEEE*

***Abstract*—In this paper, we propose a geometric approach to the theory of evidence based on convex geometric interpretations of its two key notions of belief function (b.f.) and Dempster's sum. On one side, we analyze the geometry of b.f.'s as points of a polytope in the Cartesian space called belief space, and discuss the intimate relationship between basic probability assignment and convex combination. On the other side, we study the global geometry of Dempster's rule by describing its action on those convex combinations. By proving that Dempster's sum and convex closure commute, we are able to depict the geometric structure of conditional subspaces, i.e., sets of b.f.'s conditioned by a given function *b*. Natural applications of these geometric methods to classical problems such as probabilistic approximation and canonical decomposition are outlined.**

***Index Terms*—Belief function (b.f.), belief space, conditional subspace, Dempster's rule, simplex, theory of evidence (ToE).**

I. Introduction

The theory of evidence (ToE) [1] was introduced in the late 1970s by G. Shafer as a way of representing epistemic knowledge, starting from a sequence of seminal works [2]–[4] of A. Dempster. In this formalism, the best representation of chance is a *belief function* (b.f.) rather than a Bayesian mass distribution. The b.f.'s can be pooled by means of an operator called *Dempster's rule* [2] whose appeal has made the ToE one of the most popular theories of probable reasoning. The literature on the ToE is now vast, and includes applications to fields as different as computer vision [5], social sciences [6], risk analysis [7], and sensor fusion [8]. Recent studies include, among others, the design of classifiers based on b.f.'s [9], the analysis of k-additive b.f.'s [10], and the extension of the evidential formalism to continuous spaces [11]. Those very applications stimulate, in turn, major advances in the theory itself. In estimation problems, for instance, it is often required to compute a pointwise estimate of the quantity of interest: object tracking [12] is a typical example. The problem of approximating a b.f. with a probability then naturally arises [13]–[21]. The link between b.f.'s and probabilities is as well the foundation of a popular approach to the ToE, Smets' "transferable belief model" [22].

The approximation problem, though, can be cast in a different light by asking in which space b.f.'s live, and what sort of function is the most suitable to measure distances between b.f.'s or between b.f.'s and probabilities. As a belief, function $b : 2^\Theta \to [0, 1]$ on Θ is completely specified by its $N - 1$, $N = 2^{|\Theta|}$ belief values $\{b(A), \forall A \subset \Theta, A \neq \emptyset\}$, b can be thought of as a vector $v = [v_A = b(A), A \subset \Theta, A \neq \emptyset]'$ of $\mathbb{R}^{N-1}$. The collection $\mathcal{B}$ of all points of $\mathbb{R}^N$ that correspond to a b.f. turns out to be a polytope, which we call the *belief space*.

The study of the interplay between b.f.'s and probabilities has, in fact, been posed in a geometric setup by other authors [23]–[25]. In robust Bayesian statistics, more in general, a large amount of literature exists on the study of convex sets of distributions [26]–[30].

In this paper, we introduce a geometric interpretation of the ToE, in which issues such as the probabilistic approximation problem or the description of *conditional* b.f.'s can be formalized and solved. As a reflection of the structure of the ToE, the approach is based on two pillars: the study the geometry of b.f.'s and that of Dempster's rule of combination. After recalling the basic notions of the ToE (Section II), we briefly present an example of the applications that originally motivated this work, and lay out a research plan in which geometric interpretations of b.f.'s and rule of combination are investigated (Section III). Accordingly, starting from the insight provided by the simple case of a binary frame, we discuss the convexity of the belief space and its regions associated with Bayesian and simple support b.f.'s. The first part of the paper culminates in Section IV, where we prove that $\mathcal{B}$ has the form of a *simplex*, in which the basic probability assignment [1] of a b.f. b plays the role of its simplicial coordinates in $\mathcal{B}$. The second part is devoted to the geometry of Dempster's rule. In Section VI, we prove a fundamental result on Dempster's sums of convex combinations, and use it to show that the rule of combination commutes with the convex closure operator in the belief space. This allows us to describe the "global" geometry of the orthogonal sum in terms of simplices called *conditional subspaces*. We conclude (Section VII) by giving a flavor of some of the manifold lines of research opened by the geometric approach.

II. Theory of Evidence

Definition 1: A *basic probability assignment* (b.p.a.) over a finite set (*frame of discernment* [1]) Θ is a function $m : 2^\Theta \to [0, 1]$ on its power set $2^\Theta \doteq \{A \subseteq \Theta\}$ such that $m(\emptyset) = 0$, $\sum_{A\subseteq\Theta} m(A) = 1$, and $m(A) \geq 0$ $\forall A \subseteq \Theta$. Subsets of Θ associated with nonzero values of m are called *focal elements* (f.e.'s), and their union $\mathcal{C}$ *core*. The *b.f.* $b : 2^\Theta \to [0, 1]$ associated with a b.p.a. m_b on Θ is defined as: $b(A) = \sum_{B\subseteq A} m_b(B)$.

Alternative definitions of b.f.'s can, though, be given independently from the notion of b.p.a. [1].

A b.f. $b : 2^\Theta \to [0, 1]$ is called *simple support function* focused on A whenever $m_b(A) = \sigma$, $m_b(\Theta) = 1 - \sigma$, while $m_b(B) = 0$ for every other $B \subsetneq \Theta$. On the other side, in the ToE, a finite probability function on Θ is simply a special b.f. (*Bayesian* b.f.), which assigns nonzero mass to elements of the frame only: $m_b(A) = 0$ $\forall A : |A| > 1$ (where $|A|$ denotes the

Manuscript received January 11, 2005; revised August 29, 2006 and July 2, 2007. This paper was recommended by Associate Editor A.-O. Nii.

The author is with the Institut National de Recherche en Informatique et en Automatique, Rhône-Alpes, Grenoble 38334, France (e-mail: Fabio.Cuzzolin@inrialpes.fr).

Digital Object Identifier 10.1109/TSMCC.2008.919174

cardinality of the subset A). The b.f.'s admit the order relation

$$b \leq b' \equiv b(A) \leq b'(A) \qquad \forall A \subseteq \Theta \tag{1}$$

called *weak inclusion*. A probability distribution p such that a b.f. b is weakly included in p ($p(A) \geq b(A)\ \forall A$) is said to be *consistent* with b [31]: b can be viewed as the lower envelope of the set of probabilities consistent with it: $\{p : p(A) \geq b(A)\ \forall A \subseteq \Theta\}$.

The b.f.'s representing distinct bodies of evidence can be combined through *Dempster's rule*.

Definition 2: The *orthogonal sum* or *Dempster's sum* of two b.f.'s b_1 and b_2 defined on a frame Θ is a new b.f. $b_1 \oplus b_2$ on Θ whose focal elements are all the possible nonempty intersections $A_i \cap B_j$ of f.e.'s of b_1 and b_2, respectively, and whose b.p.a. is given by

$$m_{b_1 \oplus b_2}(A) = \frac{\sum_{i,j:A_i \cap B_j = A} m_{b_1}(A_i) m_{b_2}(B_j)}{\sum_{i,j:A_i \cap B_j \neq \emptyset} m_{b_1}(A_i) m_{b_2}(B_j)} \tag{2}$$

where m_{b_1} and m_{b_2} denote the b.p.a.'s of b_1 and b_2, respectively.

We denote by $k(b_1, b_2)$ the denominator of (2). When $k(b_1, b_2) = 0$, the two b.f.'s cannot be combined. Dempster's rule can be naturally extended to the combination of several b.f.'s.

III. Geometric Approach to the Theory of Evidence

When one tries to apply the ToE to classical engineering or artificial intelligence (AI) problems, important questions arise, stimulating major advances in the theory itself. Object tracking [32], [33], for instance, is a central problem in computer vision. It concerns the reconstruction of the configuration or "pose" of a moving object (expressed as a point q of some region $\mathcal{Q}$ of $\mathbb{R}^D$, called *configuration space*) by processing the sequence of images taken during its motion. In the case of rigid bodies, their pose is simply the position and orientation of the object with respect to some fixed reference frame. If the body is "articulated" (composed by several rigid bodies) like a human arm or hand, its pose has also to describe its internal configuration.

When no *a priori* information about the body is available, the only way of doing inference on the object pose is building in a learning stage a map between poses and some salient image measurements called *features*. In [12], we proposed a method to learn those maps between a finite approximation $\tilde{\mathcal{Q}} = \{q_k, k = 1, \ldots, T\}$ of the parameter space (acquired as a collection of poses assumed by the object in a training session) and a number of feature spaces Θ_i. These maps ρ_i, together with feature Θ_i and parameter $\tilde{\mathcal{Q}}$ spaces, form what we can call an *evidential model* of the object (see Fig. 1). When the object evolves freely, the evidential model can be used to estimate its pose by representing new features as b.f.'s defined on the frames $\{\Theta_i, i\}$, projecting them onto $\tilde{\mathcal{Q}}$, and combining them through Dempster's rule. This yields a belief estimate $\hat{b} : 2^{\tilde{\mathcal{Q}}} \rightarrow [0, 1]$ of the pose, which then needs to be processed to extract a pointwise estimate $\hat{q}$ of the configuration. A natural way is to approximate $\hat{b}$ with a finite probability $\hat{p}$ on $\tilde{\mathcal{Q}}$, and later, compute its mean value as $\hat{q} = \sum_{k=1}^{T} \hat{p}(q_k) q_k$. An evidential solution to the object tracking problem involves facing the probabilistic approximation problem.

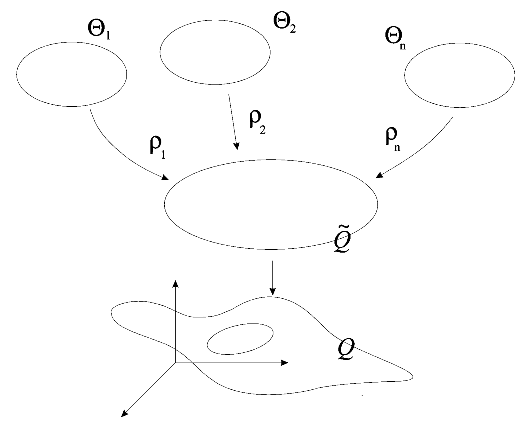

Fig. 1. Evidential model architecture.

Although the problem has been widely studied before [17] its concrete application stimulates us to pose it from scratch in a completely different setting. Where do b.f.'s live? Which relationship have they with probabilities in this space? How do you measure the distance between a b.f. and a probability? As we are going to show here, the language of convex geometry can be used to define a framework in which all those questions are addressed.

The first pillar of the ToE is the notion of basic probability assignment, i.e., the idea of assigning masses directly to events instead of elements of a frame. We then first need to understand how to describe b.p.a.'s in a convex geometric language. This leads us to define the notion of *belief space* $\mathcal{B}$ as the space of all b.f.'s on a given frame, drawing intuition from the simplest case of a binary domain. In particular, we will observe that all probabilities live in a region, which dominates the belief space (in the sense of Section II). The latter turns out to be convex, mirroring similar results for lower provisions [34]. After noticing that $\mathcal{B}$ is a triangle in the binary case, we will prove and discuss the general form of the belief space as a *polytope* or *simplex*.

The b.f.'s, though, are useful only when combined in an evidence revision process. The mechanism shaping this process in the ToE is Dempster's rule. In the second part of the paper, we will study the behavior of the rule of combination in our geometric framework, and describe the notion of *conditional b.f.* in geometric terms.

IV. Space of Belief Functions

Consider a frame of discernment Θ, and introduce in the Cartesian space $\mathbb{R}^{N-1}$, where $N = 2^{|\Theta|}$ is the number of nonempty subsets of Θ, a reference frame (a set of linearly independent vectors) $\{X_A : A \subseteq \Theta, A \neq \emptyset\}$. Each vector v of $\mathbb{R}^{N-1}$ can then be expressed in terms of this base as

$$v = \sum_{A \subseteq \Theta, A \neq \emptyset} v_A X_A = [v_A, A \subseteq \Theta, A \neq \emptyset]'.$$

For instance, if $\Theta = \{x, y, z\}$, each vector has the form $v = [v_x, v_y, v_z, v_{\{x,y\}}, v_{\{x,z\}}, v_{\{y,z\}}, v_\Theta]'$.

As each b.f. b on Θ is completely specified by its belief values $b(A)$ on all the $N-1$ subsets of Θ ($\emptyset$ can be neglected as $b(\emptyset) = 0$), v is *potentially* a b.f., its component v_A measuring

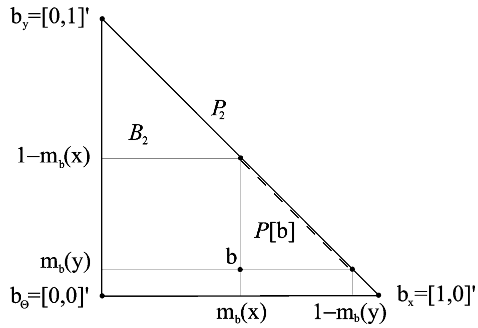

Fig. 2. The belief space $\mathcal{B}$ for a binary frame is a triangle in $\mathbb{R}^2$ whose vertices are the basis b.f.'s focused on $\{x\}, \{y\}$, and Θ (b_x, b_y, b_Θ, respectively). The probability region is the segment $Cl(b_x, b_y)$. The set of probabilities $P[b]$ consistent with a b.f. b also forms a segment.

the belief value of A: $v_A = b(A)\ \forall A \subseteq \Theta$. However, not every vector $v \in \mathbb{R}^{N-1}$ represents a valid b.f., as it may not meet the conditions of Definition 1.

Definition 3: The *belief space* is the set of points $\mathcal{B}$ of $\mathbb{R}^{N-1}$, which correspond to valid b.f.'s.

A. *Belief Space for a Binary Frame*

To get some insight about properties and geometric shape of the belief space, it may be useful to have first a look at how b.f.'s defined on a frame of discernment with just two elements $\Theta_2 = \{x, y\}$ can be represented as points of a Cartesian space. In this very simple case, each b.f. $b : 2^{\Theta_2} \to [0, 1]$ is completely determined by its belief values $b(x), b(y)$ and $b(\Theta)$ (since $b(\emptyset) = 0$ for all b). We can then collect them in a 3-D vector $[b(x), b(y), b(\Theta)]' \in \mathbb{R}^3$ and associate b with a point of $\mathbb{R}^3$. However, since it is always true that $b(\Theta) = \sum_{A\subseteq\Theta} m_b(A) = 1$, the last coordinate of the vector can also be neglected (this is, of course, true for arbitrary frames too). In the binary case, we can then represent b as the vector $[b(x) = m_b(x), b(y) = m_b(y)]'$ of $\mathbb{R}^{N-2} = \mathbb{R}^2$ (as $N = 2^2 = 4$). Since $m_b(x) \geq 0$, $m_b(y) \geq 0$, and $m_b(x) + m_b(y) \leq 1$, the set $\mathcal{B}_2$ of all possible b.f.'s on Θ_2 is the triangle in the Cartesian plane of Fig. 2, with vertices

$$b_\Theta = [0, 0]' \qquad b_x = [1, 0]' \qquad b_y = [0, 1]'$$

which correspond, respectively, to the vacuous b.f. b_Θ ($m_{b_\Theta}(\Theta) = 1$), the Bayesian b.f. b_x with $m_{b_x}(x) = 1$, and the Bayesian b.f. b_y s.t. $m_{b_y}(y) = 1$. The vectors $X_x = [1, 0]'$ and $X_y = [0, 1]'$ form a reference frame $\{X_A : \emptyset \subsetneq A \subsetneq \Theta\}$ in the Cartesian plane. All Bayesian b.f.'s on Θ_2 obey the constraint $m_b(x) + m_b(y) = 1$, and then, correspond to points of the segment $\mathcal{P}_2$ joining $b_x = [1, 0]'$ and $b_y = [0, 1]'$. Note that

$$\sum_{A\subseteq\Theta_2} b(A) = b(x) + b(y) + b(\Theta_2) = m_b(x) + m_b(y) + 1$$
$$= 2 - m_b(\Theta) \qquad (3)$$

which is equal to 2 iff b is Bayesian.

The set of Bayesian b.f.'s *consistent* with b is the segment $P[b]$ in Fig. 2 whose extreme points are the probabilities $[m_b(x), 1 - m_b(x)]'$ and $[1 - m_b(y), m_b(y)]'$. The L_1 distance between b and $P[b]$: $\|b, P[b]\|_1 = \max_{p\in P[b]}\{|b(x) - p(x)|, |b(y) - p(y)|\} = m_b(\Theta_2)$ is the mass assigned to the whole frame. From this example, we can observe the following.

1) The belief space $\mathcal{B}$ and the Bayesian space $\mathcal{P}$ are *convex*: given any two points in $\mathcal{B}$ ($\mathcal{P}$), the segment joining them is entirely in $\mathcal{B}$ ($\mathcal{P}$).

2) Moreover, $\mathcal{B}$ and $\mathcal{P}$ are both *polytopes* or *simplices*, i.e., convex closures of a finite sets of (*affinely independent*, see footnote 1) points $\mathcal{B}_2 = Cl(b_\Theta, b_x, b_y), \mathcal{P}_2 = Cl(b_x, b_y)$, where the *convex closure* of a number of vectors $v_1, \ldots, v_k$ in a Cartesian space $\mathbb{R}^m$ is defined as

$$Cl(v_1, \ldots, v_k) \doteq \left\{\sum_{i=1}^{k} \alpha_i v_i,\ \sum_{i=1}^{k} \alpha_i = 1, \alpha_i \geq 0\right\}. \qquad (4)$$

3) The probabilities consistent with a b.f. b also form a simplex (a segment, in the binary case).

B. *Region of Dominating Probabilities*

These are indeed general properties, valid for arbitrary frames. Let us first characterize the geometry of Bayesian b.f.'s. Generalizing condition (3), we can prove the following.

Theorem 1: The region of the belief space associated with all Bayesian b.f.'s on Θ is

$$\mathcal{P} = \left\{b : 2^\Theta \to [0, 1] : \sum_{A\subseteq\Theta} b(A) = 2^{n-1}\right\}, \qquad n \doteq |\Theta|. \qquad (5)$$

Proof: If $b : 2^\Theta \to [0, 1]$ is a b.f. on a frame Θ, we have that $\sum_{A\subseteq\Theta} b(A) = \sum_{A\subseteq\Theta}\sum_{B\subseteq A} m_b(B) = \sum_{B\subseteq\Theta} m_b(B)|\{A : B \subseteq A \subseteq \Theta\}|$ as each subset B is counted as many times as there are A's containing it. But, since $|\{A : B \subseteq A \subseteq \Theta\}| = |\{C \subseteq (\Theta \setminus B)\}| = 2^{|\Theta\setminus B|} = 2^{n-|B|}$, we have that (after switching back to the notation A for subsets of Θ)

$$\sum_{A\subseteq\Theta} b(A) = \sum_{A\subseteq\Theta} m_b(A)2^{n-|A|}. \qquad (6)$$

If b is Bayesian $m_b(A) = 0\ \forall A : |A| > 1$, and $\sum_{A\subseteq\Theta} b(A) = 2^{n-1}$. Conversely, if

$$\sum_{A\subseteq\Theta} b(A) = \sum_{A\subseteq\Theta} m_b(A)2^{n-|A|}$$
$$= 2^{n-1}\sum_{x\in\Theta} m_b(x) + \sum_{|A|>1} m_b(A)2^{n-|A|} = 2^{n-1}$$

then, $\sum_{x\in\Theta} m_b(x) = 1$, i.e., b is Bayesian. ■

Theorem 2: The belief space $\mathcal{B}$ is dominated by the probability region $\mathcal{P}$, namely

$$\sum_{A\subseteq\Theta} b(A) \leq 2^{n-1} \qquad \forall b \in \mathcal{B}$$

where the equality holds iff b is Bayesian.

Proof: Recalling (6), and after noticing that $2^{|\Theta\setminus B|} \leq 2^{n-1}$ (where the equality holds iff $|B| = 1$), we have that $\sum_{A\subseteq\Theta} b(A) = \sum_{A\subseteq\Theta} m_b(A)2^{|\Theta\setminus A|} \leq 2^{n-1}\sum_{A\subseteq\Theta} m_b(A) = 2^{n-1} \cdot 1$, where the equality holds iff $|A| = 1$ for every focal element of b, i.e., b is Bayesian. ■

C. Convexity

It is natural to conjecture that the belief space is convex in the general case too. In fact, it is well known that b.f.'s are a special type of *coherent lower probabilities*, (consult [34, Sec. 5.13]), and that coherent lower probabilities are closed under convex combination. This implies that convex combinations of b.f.'s are still coherent. Here, we are going to prove a stronger result. Given a b.f. b, the corresponding basic probability assignment is obtained by applying the *Möbius inversion lemma* [35]

$$m_b(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} b(B). \qquad (7)$$

We can, hence, decide whether a point $v \in \mathbb{R}^{N-1}$ is a b.f. by computing the corresponding b.p.a. and checking the axioms m_b must obey (see Definition 1). The normalization constraint $\sum_{A \subseteq \Theta} m_b(A) = 1$ trivially translates into $\mathcal{B} \subset \{v \in \mathbb{R}^{N-1} : v_\Theta = 1\}$: b.f.'s can indeed be seen as points of $\mathbb{R}^{N-2}$. The nonnegativity condition $m_b(A) \geq 0 \; \forall A \subseteq \Theta$ reads instead as

$$m_b(A) = b(A) + \cdots + (-1)^{|A|-k} \sum_{|B|=k} b(B) + \cdots + (-1)^{|A|-1} \sum_{x \in \Theta} b(x) \geq 0 \qquad \forall A \subseteq \Theta. \qquad (8)$$

All the constraints in (8) have the form $\sum_{B \subseteq A, |A \setminus B| \text{ even}} b(B) \geq \sum_{B \subseteq A, |A \setminus B| \text{ odd}} b(B)$. We can use this fact to prove the following.

Theorem 3: The belief space $\mathcal{B}$ is convex.

Proof: Let us consider two points $b_0, b_1 \in \mathcal{B}$, and prove that all the points b_α in the segment $\{b_\alpha = b_0 + \alpha(b_1 - b_0), 0 \leq \alpha \leq 1\}$ belong to $\mathcal{B}$. Since $b_0, b_1 \in \mathcal{B}$

$$\sum_{B \subseteq A, |A \setminus B| \text{ even}} b_0(B) \geq \sum_{B \subseteq A, |A \setminus B| \text{ odd}} b_0(B)$$
$$\sum_{B \subseteq A, |A \setminus B| \text{ even}} b_1(B) \geq \sum_{B \subseteq A, |A \setminus B| \text{ odd}} b_1(B) \qquad (9)$$

$\forall A \subseteq \Theta$, so that

$$\sum_{B \subseteq A, |A \setminus B| \text{ even}} b_\alpha(B) = \sum_{B \subseteq A, |A \setminus B| \text{ even}} [b_0(B) + \alpha(b_1(B) - b_0(B))]$$
$$= (1-\alpha) \sum_{B \subseteq A, |A \setminus B| \text{ even}} b_0(B) + \alpha \sum_{B \subseteq A, |A \setminus B| \text{ even}} b_1(B) \geq$$

by (9)

$$\geq (1-\alpha) \sum_{B \subseteq A, |A \setminus B| \text{ odd}} b_0(B) + \alpha \sum_{B \subseteq A, |A \setminus B| \text{ odd}} b_1(B) = \sum_{B \subseteq A, |A \setminus B| \text{ odd}} b_\alpha(B)$$

i.e., $b_\alpha \in \mathcal{B}$. ■

V. Simplicial Form of the Belief Space

It is well known that the set of probability distributions we can define on a finite sample space Θ of cardinality n forms a polytope or simplex (called *probability simplex*) in the Cartesian space $\mathbb{R}^n$, $n = |\Theta|$, whose vertices are the n versors of $\mathbb{R}^n$ itself, $[1, 0, \ldots, 0]'$, $[0, 1, \ldots, 0]'$, …, $[0, \ldots, 0, 1]'$. The belief space $\mathcal{B}$ is itself convex (Section IV-C), and corresponds to a triangle in the binary case. $\mathcal{B}$ can be indeed described as a polytope for arbitrary frames too, generalizing the case of probability distributions [23].

A. Simplex of Belief Functions

We first need to understand the geometric behavior of basic probability assignments.

Theorem 4: The set of all b.f.'s with focal elements in a collection $\{A_1, \ldots, A_m\}$ is closed and convex in $\mathcal{B}$, namely, $\{b : \mathcal{E}_b \subseteq \{A_1, \ldots, A_m\}\} = Cl(b_{A_i}, i = 1, \ldots, m)$, where $\mathcal{E}_b$ is the collection of focal elements of b, and b_A is the vector of $\mathbb{R}^{N-2}$ with components

$$b_A(B) = \begin{cases} 1, & \text{if } B \supset A \\ 0, & \text{if } B \not\supset A \end{cases} \qquad \emptyset \subsetneq B \subseteq \Theta. \qquad (10)$$

Proof: By definition, $\{b : \mathcal{E}_b \subseteq \{A_1, \ldots, A_m\}\}$ is the set of vectors of $\mathbb{R}^{N-2}$ of the form

$$b = \Big[b(A) = \sum_{B \subseteq A, B \in \mathcal{E}_b} m_b(B), \emptyset \subsetneq A \subsetneq \Theta \Big]'$$

for some collection of subsets $\mathcal{E}_b \subseteq \{A_1, \ldots, A_m\}$.

Each component $b(A)$ of these vectors b can be obviously written as $b(A) = \sum_{B \subseteq A, B \in \mathcal{E}_b} m_b(B) = \sum_{B \in \mathcal{E}_b} m_b(B) b_B(A)$ where the "indicator" function $b_B(A) = 1$ if $A \supset B$, $b_B(A) = 0$ if $A \not\supset B$ selects the subsets B of A. After collecting the values $b_B(A)$ in a vector $b_B = [b_B(A), \emptyset \subsetneq A \subsetneq \Theta]'$, we can express b as a convex combination of the vectors b_B

$$b = \sum_{B \in \mathcal{E}_b} m_b(B) b_B = \sum_{B \in \{A_1, \ldots, A_m\}} m_b(B) b_B = \sum_{i=1}^{m} m_b(A_i) b_{A_i}$$

as $m_b(B) = 0$ whenever $A_i \notin \mathcal{E}_b$. Since m_b is a basic probability assignment $\sum_{i=1}^{m} m_b(A_i) = 1$ and $m_b(A_i) \geq 0 \; \forall i$. By definition of convex closure (4)

$$\{b : \mathcal{E}_b \subseteq \{A_1, \ldots, A_m\}\} = \left\{ b = \sum_{i=1}^{m} m_b(A_i) b_{A_i}, \sum_{i=1}^{m} m_b(A_i) = 1, m_b(A_i) \geq 0 \; \forall i \right\} = Cl(b_{A_i}, i = 1, \ldots, m).$$

■

Proposition 1: The vector b_A defined by (10) is the simple support b.f. assigning unitary mass to a single subset A (Ath *basis b.f.*)

$$m_{b_A}(A) = 1 \qquad m_{b_A}(B) = 0 \qquad \forall B \neq A. \qquad (11)$$

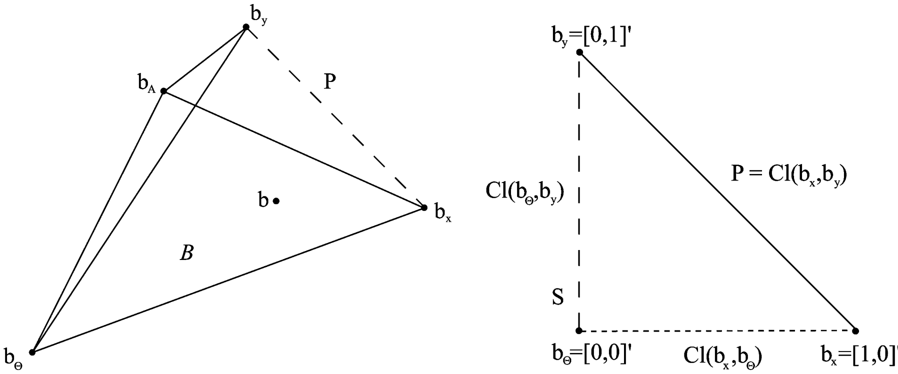

Fig. 3. (Left) Simplicial structure of the belief space $\mathcal{B} = Cl(b_A, A \neq \emptyset)$. Its vertices are all the basis b.f.'s b_A represented as vectors of $\mathbb{R}^{N-2}$. The probabilistic subspace is a subset $\mathcal{P} = Cl(b_x, x \in \Theta)$ of its border. (Right) Locations of some major classes of b.f.'s in the binary belief space.

Proof: The belief values associated with the b.p.a. (11) are $\forall B \subseteq \Theta, B \neq \emptyset$

$$b(B) = \sum_{C \subseteq B} m(C) = \begin{cases} 1, & B \supset A \\ 0, & B \not\supset A \end{cases}$$

i.e., (10). ∎

Immediately, since $\mathcal{B}$ is the collection of b.f.'s b with focal elements in $2^\Theta \setminus \emptyset$ ($\mathcal{E}_b \subseteq 2^\Theta \setminus \emptyset$), we have the following.

Corollary 1: The belief space $\mathcal{B}$ is the convex closure of all the basis b.f.'s, $\mathcal{B} = Cl(b_A, \emptyset \subsetneq A \subseteq \Theta)$.

Even though the vectors $\{b_A, \emptyset \subsetneq A \subseteq \Theta\}$ are $(N-2)$-dimensional, $\mathcal{B}$ has $N-1$ vertices, including the basis b.f. b_Θ. Since $b_\Theta(B) = 0 \ \forall B \subseteq \Theta, B \neq \Theta$, $b_\Theta = \mathbf{0}$ is the origin of $\mathbb{R}^{N-2}$.

In convex geometry, a k-dimensional *simplex* is the convex closure $Cl(x_1, \ldots, x_{k+1})$ of $k+1$ affinely independent[1] points $x_1, \ldots, x_{k+1}$ of the Cartesian space $\mathbb{R}^k$. The *faces* of a k-dimensional simplex are all the possible simplices generated by subsets of its vertices, i.e., $Cl(x_{j_1}, \ldots, x_{j_m})$ with $\{j_1, \ldots, j_m\} \subset \{1, \ldots, k+1\}$. As it is easy to see that the vectors $\{b_A, \emptyset \subsetneq A \subseteq \Theta\}$ are affinely independent in $\mathbb{R}^{N-1}$, it follows that $\mathcal{B}$ is a simplex in $\mathbb{R}^{N-1}$ (see Fig. 3-left). By Theorem 4, each b.f. $b \in \mathcal{B}$ can be written as

$$b = \sum_{\emptyset \subsetneq A \subseteq \Theta} m_b(A) b_A.$$

A b.p.a. is geometrically a choice of *simplicial coordinates* for b in the polytope $\mathcal{B}$.

B. Faces of $\mathcal{B}$ as Classes of Belief Functions

Obviously, a Bayesian b.f. (a finite probability) is a b.f. with focal elements in the collection of singletons: $\mathcal{E}_b = \{\{x_1\}, \ldots, \{x_n\}\}$. Immediately, by Theorem 4, we have the following.

Corollary 2: The region of the belief space that corresponds to probability functions is the part of its border determined by all the simple probabilities, i.e., the simplex[2] $\mathcal{P} = Cl(b_x, x \in \Theta)$.

$\mathcal{P}$ is then an $(n-1)$-dimensional face of $\mathcal{B}$ (whose dimension is instead $N - 2 = 2^n - 2$, as it has $2^n - 1$ vertices).

Some 1-D faces of the belief space also have an intuitive meaning in terms of belief. Consider the segments $Cl(b_\Theta, b_A)$ joining the vacuous b.f. b_Θ ($m_{b_\Theta}(\Theta) = 1$, $m_{b_\Theta}(B) = 0 \ \forall B \neq \Theta$) with the basis b.f. b_A (10). Points of $Cl(b_\Theta, b_A)$ can be written as a convex combination as $b = \alpha b_A + (1-\alpha) b_\Theta$. Since convex combinations are b.p.a.'s in $\mathcal{B}$, such a b.f. b has b.p.a. $m_b(A) = \alpha$, $m_b(\Theta) = 1 - \alpha$, i.e., b is a simple support function focused on A (Section II). The union of these segments for all events A: $\mathcal{S} = \cup_{\emptyset \subsetneq A \subsetneq \Theta} Cl(b_\Theta, b_A)$ is the region of all the simple support b.f.'s on Θ. In the binary case (see Fig. 3-right), simple support functions focused on $\{x\}$ lie on the horizontal segment $Cl(b_\Theta, b_x)$, while simple support b.f. focused on $\{y\}$ form the vertical segment $Cl(b_\Theta, b_y)$.

C. Geometry of Consistent Probabilities

We have seen in Section IV-A that the set $P[b]$ of the probability functions consistent with a given b.f. b is in the binary case a segment, i.e., a 1-D polytope. As a matter of fact, Ha *et al.* [23] proved that $P[b]$ can be expressed in the probability simplex as the sum of the polytopes associated with the focal elements A_i, $i = 1, \ldots, k$ of b, weighted by the corresponding masses, i.e.

$$P[b] = \sum_{i=1}^{k} m_b(A_i) conv(A_i) \qquad (12)$$

where $conv(A_i)$ is the convex closure of the probabilities assigning 1 to a particular element x of A_i. We can think of the basic probability $m_b(A)$ of a focal element A as a probability free to move inside A. Intuitively then, if we assign the mass of each focal element A_i to one of its points $x_i \in A_i$, we get an extremum of the region of consistent probabilities.

[1] An *affine combination* of k points $v_1, \ldots, v_k \in \mathbb{R}^m$ is a sum $\alpha_1 v_1 + \cdots + \alpha_k v_k$ whose coefficients sum to one: $\sum_i \alpha_i = 1$. The affine subspace generated by the points $v_1, \ldots, v_k \in \mathbb{R}^m$ is the set $\{v \in \mathbb{R}^m : v = \alpha_1 v_1 + \cdots + \alpha_k v_k, \sum_i \alpha_i = 1\}$. If $v_1, \ldots, v_k$ generate an affine space of dimension k, they are said to be *affinely independent*.

[2] With a harmless abuse of notation, we denote the basis b.f. associated with a singleton x by b_x instead of $b_{\{x\}}$. Accordingly, we write $m_b(x)$ instead of $m_b(\{x\})$.

Let us then find an explicit expression for (12). Given an arbitrary b.f. b with f.e.'s $A_1,\dots,A_k$, we can define for each choice of k representatives $[x_1,\dots,x_k]$, $x_i \in A_i\ \forall i$ of the f.e.'s the b.f.

$$b(x_1,\dots,x_k) \doteq \sum_{i=1}^{k} m_b(A_i) b_{x_i}. \qquad (13)$$

Theorem 5: $\mathcal{P}[b] = Cl(b(x_1,\dots,x_k), [x_1,\dots,x_k] \in A_1 \times \cdots \times A_k)$.

Proof: Starting from (12), $P[b]$ can be developed as

$$P[b] = \left\{ \sum_{i=1}^{k} m_b(A_i) \left(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \right), \sum_{j=1}^{|A_i|} \alpha_i^j = 1\ \forall i \right\}$$

$$= \left\{ m_b(A_1) \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} b_{x_1^{j_1}} + \sum_{i=2}^{k} m_b(A_i) \left(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \right), \sum_{j=1}^{|A_i|} \alpha_i^j = 1\ \forall i \right\}$$

$$= \left\{ m_b(A_1) \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} b_{x_1^{j_1}} + \sum_{i=2}^{k} \left(\sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} \right) m_b(A_i) \times \left(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \right), \sum_{j=1}^{|A_i|} \alpha_i^j = 1\ \forall i \right\}$$

$$= \left\{ \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} m_b(A_1) b_{x_1^{j_1}} + \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} \times \left[\sum_{i=2}^{k} m_b(A_i) \left(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \right) \right], \sum_{j=1}^{|A_i|} \alpha_i^j = 1\ \forall i \right\}$$

$$= \left\{ \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} \left[m_b(A_1) b_{x_1^{j_1}} + \sum_{i=2}^{k} m_b(A_i) \left(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \right) \right], \sum_{j=1}^{|A_i|} \alpha_i^j = 1\ \forall i \right\}. \qquad (14)$$

The expression inside the square brackets can be, in turn, written as

$$m_b(A_1) b_{x_1^{j_1}} + m_b(A_2) \sum_{j_2=1}^{|A_2|} \alpha_2^{j_2} b_{x_2^{j_2}} + \sum_{i=3}^{k} m_b(A_i) \left(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \right)$$

$$= \left(\sum_{j_2=1}^{|A_2|} \alpha_2^{j_2} \right) m_b(A_1) b_{x_1^{j_1}} + m_b(A_2) \sum_{j_2=1}^{|A_2|} \alpha_2^{j_2} b_{x_2^{j_2}} + \left(\sum_{j_2=1}^{|A_2|} \alpha_2^{j_2} \right) \sum_{i=3}^{k} m_b(A_i) \left(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \right)$$

$$= \sum_{j_2=1}^{|A_2|} \alpha_2^{j_2} \left[m_b(A_1) b_{x_1^{j_1}} + m_b(A_2) b_{x_2^{j_2}} + \sum_{i=3}^{k} m_b(A_i) \left(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \right) \right]$$

which replaced in (14) yields $P[b] = \{ \sum_{j_1=1}^{|A_1|} \sum_{j_2=1}^{|A_2|} \beta_{j_1 j_2} [m_b(A_1) b_{x_1^{j_1}} + m_b(A_2) b_{x_2^{j_2}} + \sum_{i=3}^{k} m_b(A_i) (\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j})] : \sum_{j_1 j_2} \beta_{j_1 j_2} = 1, \sum_{j=1}^{|A_i|} \alpha_i^j = 1\ \forall i = 3,\dots,k \}$ with $\beta_{j_1 j_2} \doteq \alpha_1^{j_1} \times \alpha_2^{j_2}$. Clearly the expression inside the square brackets has the same shape as before, so that, by induction on the number of focal elements, we have as desired. ■

Accordingly, the center of mass $\overline{\mathcal{P}}[b]$ of $\mathcal{P}[b]$ has the form

$$\frac{1}{\prod_i |A_i|} \sum_{[x_1,\dots,x_k] \in A_1 \times \cdots \times A_k} b(x_1,\dots,x_k)$$

$$= \frac{1}{\prod_i |A_i|} \sum_{[x_1,\dots,x_k] \in A_1 \times \cdots \times A_k} \sum_{i=1}^{k} m_b(A_i) b_{x_i}$$

$$= \frac{1}{\prod_i |A_i|} \sum_{x \in \mathcal{C}_b} b_x \cdot \sum_{A_j \supseteq \{x\}} m_b(A_j) \frac{\prod_i |A_i|}{|A_j|} \qquad (15)$$

(where $\mathcal{C}_b$ is the core of b) as each basis probability b_x appears in (15) with coefficient $m_b(A_j)$, a number of times $\prod_{i \neq j} |A_i|$ equal to the number of possible choices of representatives for the other focal elements of b. This, in turn, reads as

$$\sum_{x \in \mathcal{C}_b} b_x \left(\sum_{A_j \supseteq \{x\}} \frac{m_b(A_j)}{|A_j|} \right) = \sum_{x \in \Theta} b_x \left(\sum_{A \supseteq \{x\}} \frac{m_b(A)}{|A|} \right) \qquad (16)$$

(since no f.e. includes points outside the core), which is nothing but Smets' *pignistic function* [22]

$$\overline{\mathcal{P}}[b] = BetP[b] \qquad BetP[b](x) \doteq \sum_{A \supseteq \{x\}} \frac{m_b(A)}{|A|}.$$

The geometric analysis of the region of the consistent probabilities can be related to a popular technique in robust statistics, the Epsilon Contamination Model. For a fixed $0 < \epsilon < 1$ and a probability distribution P^*, the associated ϵ-contamination model is a convex class of distributions of the form $\{(1-\epsilon)P^* + \epsilon Q\}$, where Q is an arbitrary probability distribution. T. Seidenfeld proved that (for discrete domains) any ϵ-contamination model is equivalent to a b.f., 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 [36, Th. 2.10]). It is worth noticing that in this special case, P^* has the meaning of barycenter of the convex set, providing yet another interesting interpretation of (15).

VI. Geometry of Dempster's Rule

In the first part of the paper, we investigated the geometric properties of the twin notions of b.f. and basic probability assignment. We now know that b.f.'s live in a simplex in the Cartesian space $\mathbb{R}^{2^{|\Theta|}-2}$ whose vertices represent b.f.'s focused on a single event, and where b.p.a.'s can be interpreted as simplicial coordinates. We still need to understand the geometry of the other key concept of the ToE, Dempster's rule of combination, in turn, related to the notion of *conditional* b.f.

Conditional b.f.'s have been given in the past several alternative definitions by different authors [37]. Fagin and Halpern, for instance, defined a notion of *conditional belief* [38] as the lower envelope of a family of conditional probability functions, and provided a closed-form expression for it. On the other side, Spies [39] established a link between conditional events and discrete random sets. Conditional events were defined as *sets of equivalent events under the conditioning relation*. By applying to them a multivalued mapping (which induces a b.f., according to Dempster's original formulation), he gave a new definition of conditional b.f. An updating rule equivalent to the law of total probability when all beliefs are probabilities was introduced.

In [40], Slobodova described instead how conditional b.f.'s (defined as in Spies' approach) fit in the framework of valuation-based systems, while Xu and Smets [41], [42] showed how to use them to represent relations among variables as joint b.f.'s on the product space of the involved variables, and presented a propagation algorithm for such a network. Graphical belief models have been formulated and described [43], and the nature of belief propagations in evidential networks has been investigated [44].

In the following, we will call conditional b.f. $b|b'$ the combination of b with b':

$$b|b' = b \oplus b'.$$

In this form, conditional b.f.'s arise from the application of the ToE to estimation problems in which some sort of "temporal coherence" has to be enforced. *Data association* is a typical example.

A. Data Association, Conditional Belief Functions, and Total Belief

In the "data association" problem, a number of points moving in the 3-D space are tracked by one or more cameras and appear in an image sequence as unlabeled (undistinguishable) feature points, and we seek for the correspondences between points of two consecutive frames. A popular approach called *joint probabilistic data association* filter [45] is based on the implementation of a number of Kalman filters (each associated to a feature point) to predict the future position of the target. Unfortunately, when several features converge to a same small region of space, the algorithm cannot distinguish them anymore. However, when additional information is available, it can be used to help the association process. One way to do this is representing the evidence coming from Kalman filters and other available constraints on the targets' motion as b.f.'s, and combining them on the space of all possible associations between target points. For instance, if targets are known to belong to an articulated body of known topological model (an undirected graph whose edges represent rigid motion constraints), the rigid motion constraint can be exploited to improve the robustness of the estimation.

Formally, let us call the set of points of the model $\{M_j, j = 1, \ldots, N\}$, and $\{m_l^k, l = 1, \ldots, n(k)\}$ the measured feature points in the time-k image (where $n(k)$ is the number of detected targets). The data association problem consists on finding, at each time k, the correct association between points of the model and feature points $m_l^k \leftrightarrow M_j$. In the simplest case, we can assume $n(k) = N$. The information carried by Kalman filters' predictions concerns associations between feature points belonging to *consecutive images* $m_l^{k-1} \leftrightarrow m_m^k$, rather than points of the model, and can then be represented as b.f.'s on the frame of all *feature-to-feature* associations: $\Theta_k^{k-1} \doteq \{m_l^{k-1} \leftrightarrow m_m^k, \ \forall l, m = 1, \ldots, N\}$. The rigid motion constraint depends, on the other hand, on the *model-measurement association* at the previous step $k-1$: those associations are collected in the frame of all *past model-feature* associations: $\Theta_M^{k-1} \doteq \{m_l^{k-1} \leftrightarrow M_j, \ \forall j, l = 1, \ldots, N\}$. At each time instant k, the desired associations $m_l^k \leftrightarrow M_j$ are instead elements of the *current model-feature* associations frame: $\Theta_M^k \doteq \{m_l^k \leftrightarrow M_j, \ \forall j, l = 1, \ldots, N\}$.

The natural place where to combine all the available evidence is then the *minimal refinement*[3] of all these frames, the *combined association* frame $\Theta \doteq \Theta_M^{k-1} \otimes \Theta_k^{k-1}$. All belief constraints must be combined on Θ and projected on the current association frame Θ_M^k by restriction, producing the best current estimate.

Now, the rigid motion constraint derived from a topological model of the body can be expressed in a *conditional* way only: in fact, to test the rigidity of the motion of two measured points at time k, we need to know the correct association between points of the model and feature points at time $k-1$. Consequently, the constraint generates an *entire set* of b.f.'s $b_i : 2^{\rho_M^{k-1}(\{a_i\})} \rightarrow [0,1]$, where a_i is the ith possible model-feature $m_l^{k-1} \leftrightarrow M_j$ association at time $k-1$, and the domains $\rho_M^{k-1}(\{a_i\})$ are the elements of the partition induced on the common refinement Θ by Θ_M^{k-1} (see Fig. 4-left).

These conditional b.f.'s b_i must be reduced to a single *total b.f.* that will be eventually pooled with the other constraints. A generalization of the total probability theorem to b.f.'s is then necessary. This reads as follows.

Theorem 6: Suppose Θ and Ω are two frames of discernment, and $\rho : 2^\Omega \rightarrow 2^\Theta$ a refining. Let $b_0 : 2^\Omega \rightarrow [0,1]$ be a b.f. defined on Ω and $\{b_1, \ldots, b_n\}$ a collection of n b.f.'s $b_i : 2^{\Theta_i} \rightarrow [0,1]$ defined on the elements Θ_i the partition $\{\Theta_1, \ldots, \Theta_n\}$ of Θ induced by the coarsening Ω. Then, there exists a b.f. $b : 2^\Theta \rightarrow [0,1]$ on Θ such that:

1) *a priori constraint*: b_0 is the restriction of b to Ω and
2) *conditional constraint*: the conditional b.f. obtained by combining b with

$$b_{\Theta_i} : m_{b_{\Theta_i}}(\Theta_i) = 1 \quad m_{b_{\Theta_i}}(A) = 0 \quad \forall A : \emptyset \subsetneq A \subsetneq \Theta_i$$

coincide with b_i for all i: $b \oplus b_{\Theta_i} = b_i \ \forall i = 1, \ldots, N$.

The hypotheses of Theorem 6 are pictorially summarized in Fig. 4-right.

In the data association problem, the *a priori* constraint is the b.f. representing the estimate of the past association $\{m_l^{k-1} \leftrightarrow$

[3] Θ is called a *common refinement* [1] of a set of frames $\Theta_1, \ldots, \Theta_N$ if there exists a map (*refining*) $\rho_i : \Theta_i \rightarrow 2^\Theta$ from each frame Θ_i to a disjoint partition of Θ: $\rho_i(x) \cap \rho_i(x') = \emptyset \, \forall x, x' \in \Theta_i$; $\bigcup_{x \in \Theta_i} \rho_i(x) = \Theta$. The smallest such frame is called the *minimal refinement* $\Theta_1 \otimes \cdots \otimes \Theta_N$ of those frames. Two b.f.'s b_1, b_2 defined on two frames Θ and Ω connected by a refining ρ are said *consistent* iff $m_{b_1}(A) = m_{b_2}(\rho(A)) \ \forall A \subseteq \Theta$ and b_1 is called the *restriction* of b_2 to Θ.

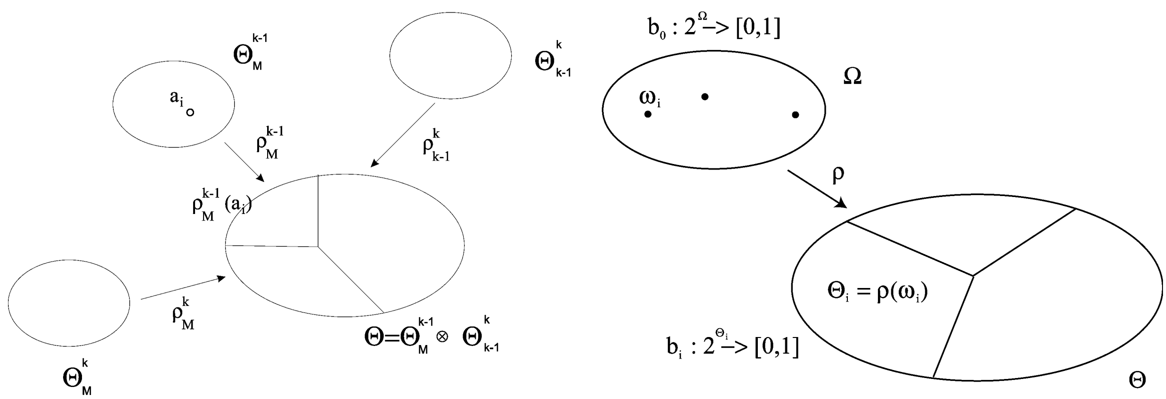

Fig. 4. (Left) Family of frames involved in the data association problem. All the constraints (expressed as b.f.'s) are combined on the common refinement Θ, and then, projected onto the current-time association frame Θ_M^k. (Right) The total belief theorem: a b.f. b on Θ such that its restriction to Ω is b_0 and whose combination with b_{Θ_i} (where $\{\Theta_1,\dots,\Theta_n\}$ is the partition of Θ induced by the refining ρ) is b_i is desired.

$M_j\}$, defined over Θ_k^{k-1} (see Fig. 4-left again). It ensures that the total b.f. is compatible with the last available estimate.

B. Dempster's Sum of Convex Combinations

The total belief theorem is only one (even though a critical one) of the theoretical issues involved by the notion of conditional b.f. In the second part of this paper, we will use the language of convex geometry that we introduced in the first part to give a characterization of the notion of conditional b.f. in the framework of the belief space. As this notion depends inherently on that of Dempster's sum, this reduces to study the geometry of the rule of combination.

We will first prove a fundamental result on Dempster's sums of convex combinations, and use it to show that the rule of combination commutes with the convex closure operator in the belief space. This will allow us to describe the "global" geometry of the orthogonal sum in terms of simplices called *conditional subspaces*, i.e., the sets of all b.f.'s conditioned by a given b.

Theorem 7: Consider a b.f. b and a collection of b.f.'s $\{b_1,\dots,b_n\}$ such that at least one of them is combinable with b. If $\sum_i \alpha_i = 1$, $\alpha_i \ge 0$ for all $i = 1,\dots,n$, then $b \oplus \sum_i \alpha_i b_i = \sum_i \beta_i (b \oplus b_i)$, where

$$\beta_i = \frac{\alpha_i k(b, b_i)}{\sum_{j=1}^n \alpha_j k(b, b_j)} \tag{17}$$

and $k(b, b_i)$ is the normalization factor for the sum $b \oplus b_i$: $k(b, b_i) \doteq \sum_{A \cap B \neq \emptyset} m_b(A) m_{b_i}(B)$.

Proof: We just need to check the equality of the corresponding basic probability assignments. After denoting by $\{B_k, k\}$, the focal elements of b_i and with $\{A_j, j\}$ those of b, the convex combination $\sum_i \beta_i b \oplus b_i$ has b.p.a. $m_{\sum_i \beta_i b \oplus b_i}(A)$ equal to

$$\sum_{B \subseteq A} (-1)^{|A-B|} \sum_i \beta_i b \oplus b_i(B)$$
$$= \sum_i \beta_i \sum_{B \subseteq A} (-1)^{|A-B|} b \oplus b_i(B) = \sum_i \beta_i m_{b \oplus b_i}(A)$$

by Moebius inversion. On the other side, by hypothesis

$$m_{\sum_i \alpha_i b_i}(A) = \sum_{B \subseteq A} (-1)^{|A-B|} \sum_i \alpha_i b_i(B)$$
$$= \sum_i \alpha_i \left(\sum_{B \subseteq A} (-1)^{|A-B|} b_i(B) \right) = \sum_i \alpha_i m_{b_i}(A).$$

Hence, after calling $E_1,\dots,E_n$ the focal elements of $\sum_i \alpha_i b_i$, we get

$$m_{b \oplus \sum_i \alpha_i b_i}(A) = \frac{\sum_{E_k \cap A_j = A} m_{\sum_i \alpha_i b_i}(E_k) m_b(A_j)}{\sum_{E_k \cap A_j \neq \emptyset} m_{\sum_i \alpha_i b_i}(E_k) m_b(A_j)}$$
$$= \frac{\sum_{E_k \cap A_j = A} \left(\sum_i \alpha_i m_{b_i}(E_k)\right) m_b(A_j)}{\sum_{E_k \cap A_j \neq \emptyset} \left(\sum_i \alpha_i m_{b_i}(E_k)\right) m_b(A_j)}$$
$$= \frac{\sum_i \alpha_i \left(\sum_{E_k \cap A_j = A} m_{b_i}(E_k) m_b(A_j)\right)}{\sum_i \alpha_i \left(\sum_{E_k \cap A_j \neq \emptyset} m_{b_i}(E_k) m_b(A_j)\right)}$$
$$= \frac{\sum_i \alpha_i \left(\sum_{B_k \cap A_j = A} m_{b_i}(B_k) m_b(A_j)\right)}{\sum_i \alpha_i \left(\sum_{B_k \cap A_j \neq \emptyset} m_{b_i}(B_k) m_b(A_j)\right)} \tag{18}$$

where the last passage holds because $m_{b_i}(E_k) = 0$ for $E_k \notin \mathcal{E}_{b_i}$, and we are left for each addenda i with the focal elements $B_k \in \mathcal{E}_{b_i}$ of b_i. Finally, we just need to note that

$$m_{b \oplus b_i}(A) = \frac{\sum_{B_k \cap A_j = A} m_{b_i}(B_k) m_b(A_j)}{\sum_{B_k \cap A_j \neq \emptyset} m_{b_i}(B_k) m_b(A_j)} \doteq \frac{N_i(A)}{k(b, b_i)}.$$

Plugging this expression in (18), we get

$$m_{b \oplus \sum_i \alpha_i b_i}(A) = \frac{\sum_i \alpha_i N_i(A)}{\sum_j \alpha_j k(b, b_j)} = \frac{\sum_i \alpha_i k(b, b_i) m_{b \oplus b_i}(A)}{\sum_j \alpha_j k(b, b_j)}$$
$$= \sum_i \beta_i m_{b \oplus b_i}(A)$$

with β_i given by (17). Now, if there is a b.f. b_j in the collection $\{b_1,\dots,b_n\}$, which is combinable with b, then $k(b, b_j) \neq 0$ and the denominator of the previous equation is nonzero, i.e., $m_{b \oplus \sum_i \alpha_i b_i}$ is well defined. ∎

Note that since $\sum_i \beta_i = 1$, $\beta_i \geq 0\ \forall i$, the combination of b with any convex sum of b.f.'s is still a convex sum of all partial combinations.

As an example, let us consider three b.f.'s in the binary frame b, b_1, b_2 with b.p.a.'s

$$m_b(x) = 1 \quad m_{b_1}(x) = 0.7 \quad m_{b_1}(\Theta) = 0.3 \quad m_{b_2}(y) = 1.$$

If we take the following convex combination $\alpha b_1 + (1-\alpha) b_2$, $\alpha = 0.6$, we have that

$$m_{\alpha b_1+(1-\alpha)b_2}(x) = 0.42 \qquad m_{\alpha b_1+(1-\alpha)b_2}(y) = 0.4$$
$$m_{\alpha b_1+(1-\alpha)b_2}(\Theta) = 0.18$$

and its combination with b yields

$$m_{b\oplus(\alpha b_1+(1-\alpha)b_2)}(x) = 1 \qquad m_{b\oplus(\alpha b_1+(1-\alpha)b_2)}(y) = 0$$
$$m_{b\oplus(\alpha b_1+(1-\alpha)b_2)}(\Theta) = 0$$

i.e., $b \oplus (\alpha b_1 + (1-\alpha) b_2) = b_x$ (the Bayesian b.f. focused on $\{x\}$).

On the other side, Theorem 7 claims that

$$b \oplus (\alpha b_1 + (1-\alpha) b_2) = \beta_1 (b \oplus b_1) + \beta_2 (b \oplus b_2)$$

with $\beta_1 = \alpha k(b, b_1)/[\alpha k(b, b_1) + (1-\alpha) k(b, b_2)]$, $\beta_2 = (1-\alpha) k(b, b_2)/[\alpha k(b, b_1) + (1-\alpha) k(b, b_2)]$, where $k(b, b_1) = 1$ and $k(b, b_2) = 0$ (as b and b_2 are not combinable), so that $\beta_1 = 1$ and $\beta_2 = 0$ and

$$b \oplus (\alpha b_1 + (1-\alpha) b_2) = b \oplus b_1 = b_1 = b_x.$$

C. Commutativity of Convex and Dempster's Combinations

In the geometric approach to the ToE, convex combinations are the geometric counterparts of basic probability assignments (Section V). Convex closure and Dempster's sum are then the two major operators acting on b.f.'s as points of the belief space. They are, in fact, inherently related to each other, as they *commute*, i.e., the order of their action on a set of b.f.'s can be exchanged. We just need to pay some attention to the issue of combinability.

Theorem 8: $b \oplus Cl(b_1, \ldots, b_k) = Cl(b \oplus b_{i_1}, \ldots, b \oplus b_{i_m})$, where $\{b_{i_1}, \ldots, b_{i_m}\} \subseteq \{b_1, \ldots, b_k\}$ are all the b.f.'s in the collection $\{b_1, \ldots, b_k\}$, which are combinable with b.

Proof: Sufficiency. We need to prove that if $b' \in b \oplus Cl(b_1, \ldots, b_k)$, then $b' \in Cl(b \oplus b_{i_1}, \ldots, b \oplus b_{i_m})$. If $b' = b \oplus \sum_{i=1}^k \alpha_i b_i$, $\sum_i \alpha_i = 1$, $\alpha_i \geq 0$, then (by Theorem 7) $b' = \sum_i \beta_i b \oplus b_i$ with β_i given by (17). But, we know that $\beta_i = 0$ iff $\nexists\, b \oplus b_i$, so that

$$b' = \sum_{i:\exists b\oplus b_i} \beta_i (b \oplus b_i) \in Cl(b \oplus b_i : \exists b \oplus b_i).$$

Necessity. We have to show that if $b' \in Cl(b \oplus b_{i_1}, \ldots, b \oplus b_{i_m})$, then for each choice of $b_{j_1}, \ldots, b_{j_l}$ not combinable with b, $b' \in b \oplus Cl(b_{i_1}, \ldots, b_{i_m}, b_{j_1}, \ldots, b_{j_l})$. If $\sum_{p=1}^m \alpha_p + \sum_{q=1}^l \alpha_q = 1$

$$b' = b \oplus \left(\sum_{p=1}^m \alpha_p b_{i_p} + \sum_{q=1}^l \alpha_q b_{j_q} \right)$$
$$= \sum_{p=1}^m \alpha'_p b \oplus b_{i_p} + \sum_{q=1}^l \alpha'_q b \oplus b_{j_q} \tag{19}$$

with (after introducing the notation $k_p = k(b, b_{i_p})$, $k_q = k(b, b_{j_q})$)

$$\alpha'_p = \frac{\alpha_p k_p}{\sum_p \alpha_p k_p + \sum_q \alpha_q k_q} \qquad \alpha'_q = \frac{\alpha_q k_q}{\sum_p \alpha_p k_p + \sum_q \alpha_q k_q}$$

by (17). But, now $k_q = 0\ \forall q$ (as b_{j_q} is not combinable with b) so that $\alpha'_q = 0$ for all $q = 1, \ldots, l$, and by (19), it follows that

$$b' = \sum_{p=1}^m \beta_p (b \oplus b_{i_p}) \in Cl(b \oplus b_{i_1}, \ldots, b \oplus b_{i_m}) \tag{20}$$

with $\beta_p = \alpha_p k_p / \sum_p \alpha_p k_p$, $\sum_p \beta_p = 1$, $\beta_p \geq 0\ \forall p$. Hence, any b.f. b' of the form (20) belongs to the region $b \oplus Cl(b_{i_1}, \ldots, b_{i_m}, b_{j_1}, \ldots, b_{j_l})$ iff we can find another collection of coefficients $\{\alpha_p,\ p = 1, \ldots, m\}$ with $\sum_p \alpha_p = 1$ such that the following constraints are met:

$$\beta_p = \frac{\alpha_p k_p}{\sum_p \alpha_p k_p} \qquad \forall p = 1, \ldots, m \tag{21}$$

(i.e., $b' = b \oplus \sum_p \alpha_p b_{i_p}$). An admissible solution of the system of equations (21) is $\tilde{\alpha}_p \doteq \beta_p / k_p$ as we get $\forall p\ \beta_p = \beta_p / \sum_p \beta_p = \beta_p$ since the β_p's sum to one, and system (21) is satisfied up to the normalization constraint. We can then normalize the solution by choosing $\alpha_p = \tilde{\alpha_p} / \sum_{p'} \tilde{\alpha_{p'}} = \beta_p / k_p \sum_{p'} (\beta_{p'} / k_{p'})$ for which (21) is still met. ∎

An immediate consequence is the following.

Corollary 3: Dempster's sum $\oplus$ and convex closure $Cl(.)$ commute, i.e., if b is combinable with $b_i\ \forall i = 1, \ldots, k$, then $b \oplus Cl(b_1, \ldots, b_k) = Cl(b \oplus b_1, \ldots, b \oplus b_k)$.

D. Conditional Subspaces

As basically a linear operator on $\mathcal{B}$, Dempster's rule commutes with convex closure (Corollary 3). This is of major importance in the framework of the geometric approach, where all major classes of b.f.'s form some sort of simplex. Using the aforementioned commutativity results, we can also identify geometric counterparts of the notions of combinability and conditioning.

Definition 4: The *conditional subspace* $\langle b \rangle$ associated with a b.f. b is the set of all the b.f.'s *conditioned by* b, namely

$$\langle b \rangle \doteq \left\{ b \oplus b', \forall b' \in \mathcal{B} \text{ s.t. } \exists\, b \oplus b' \right\}. \tag{22}$$

In rough words, the conditional subspace $\langle b \rangle$ is the possible "future" of b in a process of knowledge accumulation. As new evidence becomes available in the form of a b.f. (and is pooled through Dempster's rule), we get a series of b.f.'s $b_{t_0}, b_{t_0} \oplus b_{t_1}, b_{t_0} \oplus b_{t_1} \oplus b_{t_2}, \ldots$. The conditional subspace of the current knowledge state at time t $\langle b_{t_0} \oplus \cdots \oplus b_t \rangle$ constrains the possible outcomes of the future states of belief. Since b.f.'s are not necessarily combinable, we need first to understand the geometry of the notion of combinability.

Definition 5: The *noncombinable region* $NC(b)$ associated with a b.f. b is the collection of all the b.f.'s that are not combinable with b

$$NC(b) \doteq \{b' : \nexists\, b' \oplus b\} = \{b' : k(b, b') = 0\}.$$

The results of Section V again allow us to understand the shape of this set. As a matter of fact, the noncombinable region $NC(b)$ of b is also a simplex, whose vertices are the basis b.f.'s related to subsets disjoint from the core $\mathcal{C}_b$ of b (the union of its f.e.'s).

Proposition 2: $NC(b) = Cl(b_A, A \cap \mathcal{C}_b = \emptyset)$.

Proof: It suffices to point out that $NC(b) = \{b' : \mathcal{C}_{b'} \subseteq \overline{\mathcal{C}}_b\} = \{b' : \mathcal{E}_{b'} \subseteq 2^{\bar{\mathcal{C}}_b}\}$ where $\overline{B}$ denotes the complement of a subset B of Θ. But, by Theorem 4: $\{b' : \mathcal{E}_b \subseteq 2^{\bar{\mathcal{C}}_b}\} = Cl(b_A, A \in 2^{\bar{\mathcal{C}}_b}) = Cl(b_A : A \subseteq \bar{\mathcal{C}}_b)\} = Cl(b_A : A \cap \mathcal{C}_b = \emptyset)$. ■

Using the definition of noncombinable region $NC(b)$, we can write

$$\langle b \rangle = b \oplus (\mathcal{B} \setminus NC(b)) = b \oplus \{b' : \mathcal{C}_{b'} \cap \mathcal{C}_b \neq \emptyset\}$$

where $\setminus$ denotes the set-theoretic difference $A \setminus B = A \cap \overline{B}$. Unfortunately, $\mathcal{B} \setminus NC(b)$ does not satisfy Theorem 4. The collection of b.f.'s in $\mathcal{B} \setminus NC(b)$ cannot be written as a set of b.f.s with focal elements in a certain list, for instance $\{b' : \forall A \in \mathcal{E}_{b'}$ s.t. $A \cap \mathcal{C}_b = \emptyset\}$. In fact, a b.f. b' is combinable with b ($b' \in \mathcal{B} \setminus NC(b)$) iff *one* of its focal elements has nonempty intersection with $\mathcal{C}_b$, regardless the behavior of the others. Geometrically, this means that $\mathcal{B} \setminus NC(b)$ *is not a simplex*. Therefore, we cannot apply the commutativity results of Section VI-C directly to $\mathcal{B} \setminus NC(b)$ to find the shape of the conditional subspace. Fortunately, $\langle b \rangle$ can indeed be expressed as a Dempster's sum of b and a polytope.

Definition 6: The *compatible simplex* $C(b)$ associated with a b.f. b is the collection of all b.f.'s whose focal elements are in the core of b

$$C(b) \doteq \{b' : \mathcal{C}_{b'} \subseteq \mathcal{C}_b\} = \{b' : \mathcal{E}_{b'} \subseteq 2^{\mathcal{C}_b}\}.$$

Now, from Theorem 4, we have the following.

Corollary 4: $C(b) = Cl(b_A : A \subseteq \mathcal{C}_b)$.

The compatible simplex $C(b)$ is only a *proper* subset of the collection of b.f.'s combinable with b, $\mathcal{B} \setminus NC(b)$: nevertheless, *it contains all the relevant information*. As a matter of fact, we have the following.

Theorem 9: $\langle b \rangle = b \oplus C(b)$.

Proof: Let us denote by $\mathcal{E}_b = \{A_i, i\}$ and $\mathcal{E}_{b'} = \{B_j, j\}$ the lists of focal elements of b and b', respectively. By definition, $A_i = A_i \cap \mathcal{C}_b$ so that $B_j \cap A_i = B_j \cap (A_i \cap \mathcal{C}_b) = (B_j \cap \mathcal{C}_b) \cap A_i$, and once defined a new b.f. b'' with focal elements $\{B'_k,\ k = 1, \ldots, m\} \doteq \{B_j \cap \mathcal{C}_b,\ j = 1, \ldots, |\mathcal{E}_{b'}|\}$ (note that $m \leq |\mathcal{E}_{b'}|$ since some intersections may coincide) and basic probability assignment

$$m_{b''}(B'_k) = \sum_{j : B_j \cap \mathcal{C}_b = B'_k} m_{b'}(B'_j)$$

we have that $b \oplus b' = b \oplus b''$. ■

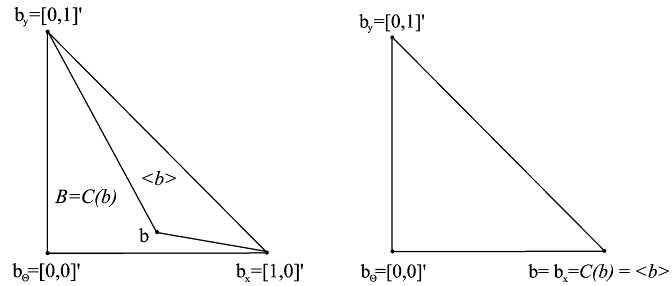

Fig. 5. Conditional subspace $\langle b \rangle$ of a b.f. b in the binary belief space $\mathcal{B}_2$, along with its compatible subspace. On the abscissa we have the belief value $b(x)$ of x, while $b(y)$ is the coordinate of b on the y axis. (Left) If b is not a basis probability its combinable simplex is $\mathcal{B}$ itself, and its conditional subspace is the triangle $Cl(b, b_x, b_y)$. (Right) If b is a basis probability, for instance if $b = b_x$, then the conditional subspace reduces to a single point.

An analogous result can be found in [1]. We are now ready to understand the convex geometry of conditional subspaces. From Theorems 4 and 9, it follows that:

Corollary 5: $\langle b \rangle = Cl(b \oplus b_A, \forall A \subseteq \mathcal{C}_b)$.

Note that, since $b \oplus b_{\mathcal{C}_b} = b$ (where $b_{\mathcal{C}_b}$ is the basis b.f. focused on the core of b), b is always one of the vertices of $\langle b \rangle$. Furthermore, $\langle b \rangle \subseteq C(b)$, since the core of a b.f. b is such that [1] $\mathcal{C}_{b \oplus b'} = \mathcal{C}_b \cap \mathcal{C}_{b'} \subseteq \mathcal{C}_b$.

1) Example: Binary Frame: Fig. 5 shows the actual shape of a conditional subspace for a b.f. defined on the simplest (binary) frame $\Theta_2 = \{x, y\}$. For each b.f. $b \in \mathcal{B}_2$, $b \neq b_x, b_y$, the noncombinable subspace is empty $NC(b) = \emptyset$, while the compatible subspace coincides with the entire belief space $C(b) = \mathcal{B}_2$ since the core of b is Θ itself. The vertices of the conditional subspace $\langle b \rangle$ are then $b \oplus b_\Theta = b$, $b \oplus b_x = b_x$, $b \oplus b_y = b_y$, and $\langle b \rangle$ is the simplex depicted in Fig. 5-left.

A singular case takes place when $b = b_x$ or $b = b_y$ (b is a basis probability assigning mass 1 to $\{x\}$ or $\{y\}$). In that case, $\mathcal{C}_b = \{x\}$ or $\{y\}$, respectively, so that $NC(b) = \{b_y\}, \{b_x\}$ and $C(b) = \{b_x\}, \{b_y\}$, respectively in the two cases. Note that, for instance, when $b = b_x$, the compatible simplex $C(b) = \{b_x\}$ is much smaller than the set of b.f.'s combinable with b, $\mathcal{B}_2 \setminus NC(b) = \mathcal{B}_2 \setminus \{b_y\}$. The conditional subspace then reduces to a single point $\langle b \rangle = \{b_x\}$ or $\{b_y\}$ (see Fig. 5-right).

VII. Applications and Developments of the Geometric Approach

To conclude, it is worth to give a flavor of the possible applications of the geometric approach to the ToE that we presented in this paper, together with a hint of the natural future developments of this framework.

A. Probabilistic Approximation of a Belief Function

One of the original motivations of this work was the probabilistic approximation problem, i.e., the question of finding the probability that is the "closest" in some sense to a given b.f. b.

1) L_1 Distance in the Belief Space: We may consider the possibility of using the L_1 norm to measure distances in the belief space: $\|b - b'\|_{L_1} \doteq \sum_{A \subseteq \Theta} |b(A) - b'(A)|$. Unfortunately, this norm is of no use as all Bayesian b.f.'s consistent with b have the same L_1 distance from b.

Lemma 1: If b dominates b', $b \geq b'$, then $\mathcal{C}_b \subseteq \mathcal{C}_{b'}$.

Proof: Obviously, since $b(A) \geq b'(A)$ for every $A \subseteq \Theta$, that is true for $\mathcal{C}_{b'} \subseteq \Theta$ too, which is also a subset of Θ. As $b'(\mathcal{C}_{b'}) = 1$ we have $1 \geq b(\mathcal{C}_{b'}) \geq 1$, i.e., $b(\mathcal{C}_{b'}) = 1$. By definition of core ($b(A) = 1$ iff $A \supseteq \mathcal{C}$), this is equivalent to $\mathcal{C}_{b'} \supseteq \mathcal{C}_b$. ■

Theorem 10: If $b : 2^\Theta \to [0,1]$ is an arbitrary b.f. on a frame Θ, then the L_1 distance between b and any Bayesian b.f. p is the same for all $p \in P[b]$ consistent with b

$$\|p - b\|_{L_1} \doteq \sum_{A \subseteq \Theta} |p(A) - b(A)| = 2^{|\Theta \setminus \mathcal{C}_b|} \left[2^{|\mathcal{C}_b|-1} - 1 - \sum_{A \subsetneq \mathcal{C}_b} b(A) \right]. \quad (23)$$

Proof: Lemma 1 guarantees that $\mathcal{C}_p \subseteq \mathcal{C}_b$, so that $p(A) - b(A) = 1 - 1 = 0$ for $A \supseteq \mathcal{C}_b$. On the other hand, if $A \cap \mathcal{C}_b = \emptyset$, then $p(A) - b(A) = 0 - 0 = 0$. We are then left with sets that correspond to unions of nonempty proper subsets of $\mathcal{C}_b$ and arbitrary subsets of $\bar{\mathcal{C}}_b = \Theta \setminus \mathcal{C}_b$. By definition, if $A' = A \cup B$ with $A \subsetneq \mathcal{C}_b, B \subseteq \bar{\mathcal{C}}_b$, we have $b(A') = b(A)$ (see Theorem 9).

For each $A \subseteq \mathcal{C}_b, A \neq \mathcal{C}_b$ there are $2^{|\Theta \setminus \mathcal{C}_b|}$ such subsets A' containing it, so that

$$\begin{aligned}\sum_{A \subseteq \Theta} |p(A) - b(A)| &= \sum_{A' = A \cup B, A \subsetneq \mathcal{C}_b, B \subseteq \bar{\mathcal{C}}_b} |p(A') - b(A')| \\ &= \sum_{A \subsetneq \mathcal{C}_b} 2^{|\Theta \setminus \mathcal{C}_b|} |p(A) - b(A)| \\ &= 2^{|\Theta \setminus \mathcal{C}_b|} \sum_{A \subsetneq \mathcal{C}_b} |p(A) - b(A)| \\ &= 2^{|\Theta \setminus \mathcal{C}_b|} \left[\sum_{A \subsetneq \mathcal{C}_b} p(A) - \sum_{A \subsetneq \mathcal{C}_b} b(A) \right]\end{aligned}$$

since $p(A) \geq b(A)\ \forall A$. But then, by (6), $\sum_{A \subsetneq \mathcal{C}_b} p(A) = 2^{|\mathcal{C}_b|-1} - p(\mathcal{C}_b) = 2^{|\mathcal{C}_b|-1} - 1$, and we get (23). ■

2) L_2 Distance and Orthogonal Projection: Since the L_1 norm is not suitable as a distance between b.f.'s, we can think of using the standard Euclidean distance $\|b - b'\|_{L_2} = \sqrt{\sum_{A \subseteq \Theta} |b(A) - b'(A)|^2}$. The Bayesian simplex $\mathcal{P}$ determines a linear subspace of $\mathbb{R}^{N-2}$. It makes sense then to define the *orthogonal projection* of a b.f. b onto $\mathcal{P}$. By definition, the orthogonal projection $\pi[b]$ of b onto $\mathcal{P}$ is the unique Bayesian function that minimizes the L_2 distance between b and $\mathcal{P}$ in the belief space

$$\pi[b] = \arg\min_{p \in \mathcal{P}} \|p - b\|_{L_2} = \arg\min_{p \in \mathcal{P}} \sqrt{\sum_{A \subseteq \Theta} |b(A) - p(A)|^2}.$$

In [46], we studied the problem of finding the expression of $\pi[b]$ in terms of the belief values of the original b.f. b, and proved a number of properties it possesses. In particular, we showed that the orthogonal projection commutes with convex combination, $\pi[\sum_i \alpha_i b_i] = \sum_i \alpha_i \pi[b_i]\ \sum_i \alpha_i = 1$ mirroring a similar property of the pignistic function.

3) Approximation Criterion Based on Dempster'S Rule: Of course, many different optimization criteria can be proposed, yielding distinct approximation problems. However, the rule of combination is central in the ToE: b.f.'s are useful only when combined with others in a reasoning process. It is natural to think that this should be taken into account when tackling the approximation problem. A possible way to comply is to formulate an optimization problem based on the "external" behavior of the desired approximation.

Criterion: A good approximation of a b.f., when combined with any other b.f., must produce results similar to those obtained by combining the original b.f.

Analytically, this translates as the following optimization problem:

$$\hat{b} = \arg\min_{b'' \in \mathcal{A}} \int_{b' \in \mathcal{B}} dist(b \oplus b', b'' \oplus b') db' \quad (24)$$

where b is the original b.f. to approximate, $b' \in \mathcal{B}$ is an arbitrary b.f. on the same frame, $dist$ is some distance function, and $\mathcal{A}$ is the class of b.f.'s that the approximation belongs to. The role of $\oplus$ can be played by any other meaningful operator, like, for instance, the disjunctive rule of combination for unnormalized b.f.'s [47]. Possibly, the resulting approximation should be independent from the choice of the distance actually used in (24). Let us consider, in particular, the class $\mathcal{A} = \mathcal{P}$ of all Bayesian b.f.'s. As the *relative plausibility of singletons* [20] $\tilde{pl}_b(x) \doteq pl_b(x) / \sum_{y \in \Theta} pl_b(y)$ (where $pl_b(A) = 1 - b(A^c)$) perfectly represents b when combined with any Bayesian b.f. ($b \oplus p = \tilde{pl}_b \oplus p\ \forall p \in \mathcal{P}$), the modified version of (24) in which the original b.f. is combined with Bayesian b.f.'s only is trivially solved by $\tilde{pl}_b$

$$\tilde{pl}_b = \arg\min_{p \in \mathcal{P}} \int_{p' \in \mathcal{P}} \|b \oplus p' - p \oplus p'\| dp' \quad (25)$$

whatever the norm we choose, as $b \oplus p' - \tilde{pl}_b \oplus p' = 0\ \forall p'$. It is then natural to conjecture that the relative plausibility function could be the solution of the general approximation problem (24), too. We will work on this conjecture in the near future.

B. Geometry of Possibility Measures

Consonant b.f.'s are b.f.'s whose focal elements are nested: they are the counterparts in the ToE of *possibility measures* [48]. All possible lists of f.e.'s associated with consonant b.f.'s then correspond to all possible chains of subsets $A_1 \subset \cdots \subset A_m$ of Θ. Theorem 4 implies that all b.f.'s whose focal elements belong to a chain $X = \{A_1, \ldots, A_m\}$ form the simplex $Cl(b_{A_1}, \ldots, b_{A_m})$. The region of the belief space formed by consonant b.f.'s is then the union of a collection of convex components, each associated with a different *maximal* chain $\mathcal{A}$

$$\mathcal{CO} = \bigcup_{\mathcal{A} = A_1 \subset \ldots \subset A_n} Cl(b_{A_1}, \ldots, b_{A_n}).$$

The number of convex components of $\mathcal{CO}$ is then the number of maximal chains in 2^Θ, i.e., $\prod_{k=1}^n \binom{k}{1} = n!$. Since the length of a maximal chain is the cardinality n of Θ, the dimension of these convex components is $\dim Cl(b_{A_1}, \ldots, b_{A_n}) = n - 1$.

In [49], we showed that $\mathcal{CO}$ has the form of a *simplicial complex*, i.e., a collection of simplices such that: 1) if a simplex belongs to the collection, then all its faces of any dimension also belong to it and 2) the intersection of two simplices is a face of both. The geometric description of consonant b.f.'s pictures then a sort of duality between probability and possibility measures, represented by the dichotomy simplex—simplicial complex. It is not hard to show that this is due to the special relation between those measures and the norms L_1 and L_∞, respectively, as probability and possibility of an event A are

$$P(A) = \sum_{x \in A} P(x) \qquad Pos(A) = \max_{x \in A} Pos(x).$$

Recalling Section VII-A, the duality principle would then imply to choose as possibilistic approximation (see also [50] and [51]) of a b.f. b, according to the optimization criterion (24), the unique consonant b.f. c with plausibility

$$pl_c(A) = \frac{\max_{x \in A} pl_b(x)}{\max_{x \in \Theta} pl_b(x)}. \tag{26}$$

A formal proof of this conjecture will be object of future work.

C. Canonical Decomposition

A *separable support function* is a b.f. that is either a simple support b.f., or is equal to the orthogonal sum of two or more simple support functions, namely, $b = b_1 \oplus \cdots \oplus b_n$ where $n \geq 1$, and b_i is a simple support b.f. $\forall i = 1, \ldots, n$. Separable support functions can be decomposed in different ways. However [1], we have the following.

Proposition 3: If $b \neq b_\Theta$ is a nonvacuous separable support function with core $\mathcal{C}_b$, then there exists a *unique* collection $b_1, \ldots, b_n$ of nonvacuous simple support functions satisfying the following conditions: 1) $n \geq 1$; 2) $b = b_1$ if $n = 1$, and $b = b_1 \oplus \cdots \oplus b_n$ if $n \geq 1$; 3) $\mathcal{C}_{b_i} \subseteq \mathcal{C}_b$; and 4) $\mathcal{C}_{b_i} \neq \mathcal{C}_{b_j}$ if $i \neq j$.

This unique decomposition is called *canonical decomposition*. Smets [52] and Kramosil [53] solved the canonical decomposition problem by means of algebraic and measure-theoretic methods, respectively. Schubert [54] has also studied the issue. We can nevertheless think of using our knowledge of the shape of conditional subspaces (Theorem 9) to find the simple components of a separable b.f. b. It is indeed quite easy to note that in the binary case ($b \in \mathcal{B}_2$), the simple components e_x, e_y of a separable support b.f. can be expressed as $e_x = Cl(b, b \oplus b_y) \cap Cl(b_\Theta, b_x) = Cl(b, b_y) \cap Cl(b_\Theta, b_x)$, $e_y = Cl(b, b_x) \cap Cl(b_\Theta, b_y)$ (see Fig. 6), and have coordinates

$$e_x = \left[\frac{m_b(x)}{1 - m_b(y)}, 0\right]' \qquad e_y = \left[0, \frac{m_b(y)}{1 - m_b(x)}\right]'.$$

A general geometric proof of the solution looks then well within reach.

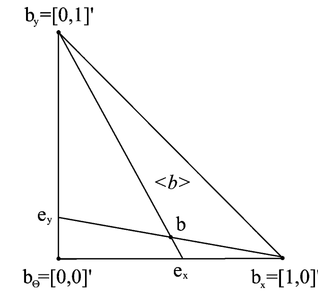

Fig. 6. Canonical decomposition of a separable support b.f. in the binary belief space.

VIII. Conclusion and Perspectives

In this paper, we introduced a geometric approach to the ToE in which b.f.'s are thought of as points of a Cartesian space. Starting from the insight provided by the binary case, we proved the convexity of the belief space, and showed that $\mathcal{B}$ has, in fact, the form of a simplex. We proved an important result on Dempster's sums of convex combinations, and used it as a tool to show that the rule of combination commutes with the convex closure operator in the belief space. This finally allowed us to describe the "global" geometry of the orthogonal sum in terms of simplices called conditional subspaces. Straightforward applications of the presented approach are among others the probabilistic approximation problem and the canonical decomposition of a b.f. into simple support b.f.'s, while a natural prosecution of the research is the study of the pointwise behavior of Dempster's rule. In a wider context, a description of the geometry of possibility measures or consonant b.f.'s can be seen as a first step toward a unified geometric interpretations of uncertainty measures.

References

[1] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.
[2] A. Dempster, "Upper and lower probability inferences based on a sample from a finite univariate population," *Biometrika*, vol. 54, pp. 515–528, 1967.
[3] A. Dempster, "Upper and lower probabilities generated by a random closed interval," *Ann. Math. Stat.*, vol. 39, pp. 957–966, 1968.
[4] A. Dempster, "Upper and lower probabilities inferences for families of hypothesis with monotone density ratios," *Ann. Math. Stat.*, vol. 40, pp. 953–969, 1969.
[5] B. Ristic and P. Smets, "The TBM global distance measure for the association of uncertain combat id declarations," *Inf. Fusion*, vol. 7, no. 3, pp. 276–284, 2006.
[6] K. Lo, "Agreement and stochastic independence of belief functions," *Math. Soc. Sci.*, vol. 51, no. 1, pp. 1–22, 2006.
[7] S. Demotier, W. Schon, and T. Denoeux, "Risk assessment based on weak information using belief functions: A case study in water treatment," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 36, no. 3, pp. 382–396, May 2006.
[8] D. Mercier, T. Denoeux, and M. Masson, "Refined sensor tuning in the belief function framework using contextual discounting," in *Proc. IPMU*, 2006, pp. 1443–1450.
[9] B. Quost, T. Denoeux, and M. Masson, "One-against-all classifier combination in the framework of belief functions," presented at the IPMU, Paris, France, 2006.
[10] P. Miranda, M. Grabisch, and P. Gil, "On some results of the set of dominating k-additive belief functions," in *Proc. IPMU*, 2004, pp. 625–632.
[11] B. Ristic and P. Smets, "Belief function theory on the continuous space with an application to model based classification," in *Proc. IPMU*, 2004, pp. 1119–1126.

[12] F. Cuzzolin and R. Frezza, "Evidential modeling for pose estimation," presented at the 4rd Int. Symp. Imprecise Probabilities Their Appl. (ISIPTA'05), Pittsburgh, PA.
[13] A. B. Yaghlane, T. Denoeux, and K. Mellouli, "Coarsening approximations of belief functions," in *Proc. ECSQARU'2001*, 2008, S. Benferhat and P. Besnard, Eds., pp. 362–373.
[14] T. Denoeux, "Inner and outer approximation of belief structures using a hierarchical clustering approach," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 9, no. 4, pp. 437–460, 2001.
[15] T. Denoeux and A. B. Yaghlane, "Approximating the combination of belief functions using the Fast Moebius Transform in a coarsened frame," *Int. J. Approx. Reasoning*, vol. 31, no. 1-2, pp. 77–101, Oct. 2002.
[16] R. Haenni and N. Lehmann, "Resource bounded and anytime approximation of belief function computations," *Int. J. Approx. Reasoning*, vol. 31, no. 1-2, pp. 103–154, Oct. 2002.
[17] M. Bauer, "Approximation algorithms and decision making in the Dempster-Shafer theory of evidence—An empirical study," *Int. J. Approx. Reasoning*, vol. 17, pp. 217–237, 1997.
[18] B. Tessem, "Approximations for efficient computation in the theory of evidence," *Artif. Intell.*, vol. 61, no. 2, pp. 315–329, 1993.
[19] J. Lowrance, T. Garvey, and T. Strat, "A framework for evidential-reasoning systems," in *Proc. Nat. Conf. Artif. Intell.,* A. A. for Artificial Intelligence, Ed., 1986, pp. 896–903.
[20] F. Voorbraak, "A computationally efficient approximation of Dempster–Shafer theory," *Int. J. Man-Mach. Stud.*, vol. 30, pp. 525–536, 1989.
[21] B. Cobb and P. Shenoy, "A comparison of Bayesian and belief function reasoning," *Inf. Syst. Frontiers*, vol. 5, no. 4, pp. 345–358, 2003.
[22] P. Smets, "Belief functions versus probability functions," in *Uncertainty and Intelligent Systems*, L. Saitta, B. Bouchon, and R. Yager, Eds. Berlin, Germany: Springer-Verlag, 1988, pp. 17–24.
[23] V. Ha and P. Haddawy, "Geometric foundations for interval-based probabilities," in *Proc. KR'98*, A. G. Cohn, L. Schubert, and S. C. Shapiro, Eds. San Francisco, CA: Morgan Kaufmann, 1998, pp. 582–593.
[24] P. Black, "Geometric structure of lower probabilities," in *Random Sets: Theory and Applications*, Goutsias, Malher, and Nguyen, Eds. New York: Springer-Verlag, 1997, pp. 361–383.
[25] M. Daniel, "Consistency of probabilistic transformations of belief functions," in *Proc. IPMU*, 2004, pp. 1135–1142.
[26] T. Melkonyan and R. Chambers, "Degree of imprecision: Geometric and algebraic approaches," *Int. J. Approximate Reasoning*, vol. 24, no. 2, pp. 522–538, 2006.
[27] F. Cozman, "Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions," *J. Comput. Graphical Statist.*, vol. 8, no. 4, pp. 824–838, 1999.
[28] Berger, "Robust Bayesian analysis: Sensitivity to the prior," *J. Statist. Planning Inference*, vol. 25, pp. 303–328, 1990.
[29] T. Herron, T. Seidenfeld, and L. Wasserman, "Divisive conditioning: Further results on dilation," *Philos. Sci.*, vol. 64, pp. 411–444, 1997.
[30] T. Seidenfeld and L. Wasserman, "Dilation for convex sets of probabilities," *Ann. Statist.*, vol. 21, pp. 1139–1154, 1993.
[31] H. Kyburg, "Bayesian and non-Bayesian evidential updating," *Artif. Intell.*, vol. 31, no. 3, pp. 271–294, 1987.
[32] T. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Image Vision Comput.*, vol. 81, pp. 231–268, 2001.
[33] D. Gavrila and L. Davis, "3D model-based tracking of humans in action: A multi-view approach," in *Proc. CVPR'96*, San Francisco, CA, pp. 73–80.
[34] P. Walley, *Statistical Reasoning with Imprecise Probabilities*. London, U.K.: Chapman & Hall, 1991.
[35] M. Aigner, *Combinatorial Theory (Classics in Mathematics)*. New York: Springer-Verlag, 1979.
[36] T. Seidenfeld, "Some static and dynamic aspects of rubust Bayesian theory," in *Random Sets: Theory and Application*, Goutsias, Malher, and Nguyen, Eds. New York: Springer-Verlag, 1997, pp. 385–406.
[37] G. Shafer, "Belief functions and parametric models," *J. R. Statist. Soc., Series B*, vol. 44, pp. 322–352, 1982.
[38] R. Fagin and J. Halpern, "A new approach to updating beliefs," in *Uncertainty in Artificial Intelligence,* vol. 6, L. K. P. P. Bonissone, M. Henrion, and J. Lemmer, Eds., 1991, pp. 347–374.
[39] M. Spies, "Conditional events, conditioning, and random sets," *IEEE Trans. Syst., Man, Cybern.*, vol. 24, pp. 1755–1763, 1994.
[40] A. Slobodova, "Conditional belief functions and valuation-based systems," Inst. Control Theory Robotics, Slovak Acad. Sci., Bratislava, SK, Rep., 1994.
[41] H. Xu and P. Smets, "Evidential reasoning with conditional belief functions," in *Proc. 10th Uncertainty Artif. Intell.*, L. de Mantaras R. and P. D., Eds., 1994, pp. 598–605.
[42] H. Xu and P. Smets, "Reasoning in evidential networks with conditional belief functions," *Int. J. Approx. Reasoning*, vol. 14, pp. 155–185, 1996.
[43] R. G. Almond, *Graphical Belief Modeling*. London, U.K./Boca Raton, FL: Chapman & Hall/CRC, 1995.
[44] B. B. Yaghlane and K. Mellouli, "Belief function propagation in directed evidential networks," in *Proc. IPMU*, 2006, pp. 47–70.
[45] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*. New York: Academic, 1988.
[46] F. Cuzzolin, "Two new Bayesian approximations of belief functions based on convex geometry," *IEEE Trans. Syst., Man Cybern. B, Cybern.*, vol. 37, no. 4, pp. 993–1008, Aug. 2007.
[47] P. Smets, "The nature of the unnormalized beliefs encountered in the transferable belief model," in *Proc. 8th Annu. Conf. Uncertainty Artif. Intell. (UAI-92)*. San Mateo, CA: Morgan Kaufmann, pp. 292–329.
[48] C. Joslyn, "Towards an empirical semantics of possibility through maximum uncertainty," in *Proc. IFSA 1991*, vol. A. R. Lowen and M. Roubens, Eds., pp. 86–89.
[49] F. Cuzzolin, "Simplicial complexes of finite fuzzy sets," in *Proc. 10th Int. Conf. Inf. Process. Manage. Uncertainty IPMU'04*, Perugia, Italy, pp. 1733–1740.
[50] D. Dubois and H. Prade, "Consonant approximations of belief functions," *Int. J. Approx. Reasoning*, vol. 4, pp. 419–449, 1990.
[51] P. Baroni, "Extending consonant approximations to capacities," in *Proc. IPMU*, 2004, pp. 1127–1134.
[52] P. Smets, "The canonical decomposition of a weighted belief," in *Proc. Int. Joint Conf. AI, IJCAI'95*, Montréal, QC, Canada, pp. 1896–1901.
[53] I. Kramosil, "Measure-theoretic approach to the inversion problem for belief functions," *Fuzzy Sets Syst.*, vol. 102, pp. 363–369, 1999.
[54] J. Schubert, "Managing decomposed belief functions," presented at the IPMU, Paris, France, 2006.

Fabio Cuzzolin (M'01) was born in Jesolo, Italy, in 1971. He received the Laurea degree (magna cum laude) and the Ph.D. degree on visions of a generalized probability theory from the University of Padova, Padua, Italy, in 1997 and 2001, respectively.

He was with the Autonomous Navigation and Computer Vision Laboratory (NAVLAB), University of Padova. He has been a Postodoctoral Researcher with the Politecnico di Milano, Milan, Italy, and the University of California, Los Angeles (UCLA) Vision Laboratory, Los Angeles. He is currently engaged in the Perception project at the Institut national de recherche en informatique et en automatique (INRIA) Rhone-Alpes, Grenoble, France. His current research interests include computer vision applications like gesture and action recognition, object pose estimation, and identity recognition from gait. His main field of investigation, however, remains that of generalized and imprecise probabilities. In particular, he has formulated a geometric approach to the theory of belief functions, focusing mainly on the probabilistic approximation problem, and studied the notion of independence of sources from an algebraic point of view. He collaborates with several journals, including the International Journal of Approximate Reasoning and Information Fusion.

Dr. Cuzzolin is a member of the Society for Imprecise Probabilities and Their Applications. He also collaborates with the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS, PART B.