HAL
open science

A geometric approach to the theory of evidence

Fabio Cuzzolin

To cite this version:

Fabio Cuzzolin. A geometric approach to the theory of evidence. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2008, 38 (4), pp.522-534. 10.1109/TSMCC.2008.919174 . inria-00590222

HAL Id: inria-00590222

https://inria.hal.science/inria-00590222v1

Submitted on 3 May 2011

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

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

A geometric approach to the theory of evidence

Fabio Cuzzolin

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 and Dempster's sum. On one side, we analyze the geometry of belief functions 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 become able to depict the geometric structure of conditional subspaces, i.e. sets of belief functions conditioned by a given function b. Natural applications of these geometric methods to classical problems like probabilistic approximation and canonical decomposition are outlined.

Index Terms

Theory of evidence, belief function, belief space, simplex, Dempster's rule, commutativity, conditional subspace.

I. Introduction

The *theory of evidence* (ToE) [42] was introduced in the late Seventies by Glenn Shafer as a way of representing epistemic knowledge, starting from a sequence of seminal works ([18], [19], [20]) of Arthur Dempster. In this formalism the best representation of chance is a *belief function* (b.f.) rather than a Bayesian mass distribution. Belief functions assign probability values to *sets* of possibilities rather than single events: their appeal rests on the fact they naturally encode evidence in favor to *propositions*. The theory embraces the familiar idea of assigning numbers between 0 and 1 to indicate degrees of belief but, instead of focusing on *how* these numbers are

F. Cuzzolin is with the Vision Laboratory, Computer Science Department, University of California at Los Angeles, USA. E-mail: cuzzolin@cs.ucla.edu. He is also with the Perception group, INRIA Rhône-Alpes, Grenoble, France.

determined, it concerns their combination. A simple method for combining the evidence carried by a number of different sources called *Dempster's rule* [18] is provided which makes no use of any a-priori distribution. In this sense, following Shafer, it can be seen as a theory of probable reasoning.

The literature on the ToE is now vast, and includes applications to fields like computer vision [38], social sciences [31], risk analysis [17], sensor fusion [33]. Recent studies include the design of classifiers based on belief functions [36], the analysis of k-additive b.f. [34], and the extension of the evidential formalism to continuous spaces [37].

When one tries and apply the theory of evidence to classical vision problems, however, a number of important question arises, stimulating major advances in the theory itself.

In the *object tracking* context, for instance, at each time instant an estimate of the current configuration or "pose" $q(t)$ of an articulated object, given a sequence of images of the moving body, is desired. Using an evidential approach [13] image measurements can be represented as belief functions and combined by means of Dempster's rule to produce an estimate of the pose $\hat{q}(t) \in \tilde{\mathcal{Q}}$, where $\tilde{\mathcal{Q}}$ is a "good" finite approximation of the configuration space $\mathcal{Q}$ of the object. This "belief estimate" of the pose expressed as the belief function on $\tilde{\mathcal{Q}}$ which emerges from the feature integration mechanism needs to be processed to extract a point-wise estimate $\hat{q}(t)$ of the pose. One way to do that is to compute the "best" probabilistic approximation $p : \tilde{\mathcal{Q}} \rightarrow [0, 1]$ (in some sense) of the belief estimate and computing the corresponding expected pose.

The problem of how to find such an approximation then arises. Many people have in fact worked on this problem, and a number of papers [53], [21], [22], [28] have been published on this issue (see [6], [5] for a review), mainly in order to find efficient implementations of the rule of combination aiming to reduce the number of focal elements. Tessem [48], for instance, incorporated only the highest-valued focal elements in his m_{klx} approximation; a similar approach inspired the *summarization* technique formulated by Lowrance *et al.* [32]. The connection between belief functions and probabilities is as well the basement of a popular approach to the theory of evidence, Smets' *pignistic* model [44], in which beliefs are represented at credal level (as convex sets of probabilities), while decisions are made by resorting to a Bayesian belief function called *pignistic transformation*. On his side, in his 1989 paper [49] F. Voorbraak proposed to adopt the so-called *relative plausibility* function $\tilde{pl}_b$, the unique probability that, given a belief function b with plausibility pl_b, assigns to each singleton its normalized plausibility.

Cobb and Shenoy [11], [9], [10] described the properties of this relative plausibility of singletons and discussed its nature of probability function equivalent to the original belief function when combined with another probability.

The approximation problem, however, can be cast in a different light, by asking in which space belief functions live, and what sort of distance is the most suitable to measure distances between b.f. or between b.f. and probabilities.

It is quite well-known that the set of probability distributions over a finite sample space Θ of size n can be represented as a polytope (called *probability simplex*) in the Euclidean space $\mathbb{R}^n$, whose vertices correspond to probabilities focused on a single element

$$[1, 0, ..., 0]', [0, 1, ..., 0]', \cdots, [0, ..., 0, 1]'.$$

Analogously, as a belief function $b : 2^\Theta \to [0, 1]$ on Θ is completely specified by its $N - 1$, $N = 2^{|\Theta|}$ belief values

$$\{b(A), \forall A \subset \Theta, A \neq \emptyset\}, \tag{1}$$

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 the points of $\mathbb{R}^N$ which correspond to a belief function turns out to be a polytope too, which we call *belief space*. The approximation problem and a comparative study of admissible distance functions can then naturally be posed in the framework of the belief space.

Going back to vision, another interesting task called *data association* [3] consists on reconstructing the association between clouds of moving targets appearing in consecutive images. If those points belong to an articulated body whose topological model is known (i.e. we know which pairs of targets are constrained to move rigidly), the rigid motion constraint can be used in order to achieve the desired correspondence. Different motion constraints can be represented as belief functions, and combined on the space of all the possible associations between targets of two consecutive images. However, many of those constraints can be expressed in conditional way only, the idea of combination of *conditional belief functions* in a filtering-like process has to be addressed.

In this work we introduce a geometric interpretation of the theory of evidence, in which both the probabilistic approximation problem and the description of conditional belief functions can be formalized and solved. We will study the geometry of belief functions and Dempster's rule of combination, which provide the basic tools of a geometric approach to the ToE.

The study of the interplay between belief functions and probabilities has in fact been posed in a geometric setup by other authors [26], [8], [7], [16]. P. Black, in particular, dedicated his doctoral thesis to the study of the geometry of belief functions and other monotone capacities [7]. An abstract of his results can be found in [8], 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.

Another close reference is perhaps a work [26] of Ha and Haddawy where they propose a framework in which an "affine operator" is used as a tool for constructing convex sets of probability distributions, and can be considered a generalization of both belief functions and interval probabilities. Uncertainty is modeled as a set of probabilities represented as "affine trees", while actions (modifications of the uncertain state) are defined as tree manipulators. A small number of properties of the affine operator are also presented. In a later work [27] they present the interval generalization of the probability cross-product operator called convex-closure (cc) operator. They provide an analysis of the properties of the cc-operator relative to manipulations of sets of probabilities, and formulate interval versions of Bayesian propagation algorithms based on it. Probability intervals are represented in a computationally efficient fashion, by means of a data structure called *pcc-tree*, in which branches are annotated with intervals, and nodes with convex sets of probabilities.

In this paper, after recalling the basic notions of the theory of evidence (Section II), we will briefly present the applications which originally motivated this work, and lay out a research plan in which geometric interpretations of the two pillars of the ToE (belief functions and rule of combination) are investigated (Section III). Accordingly in Section IV, starting from the insight provided by the simple case of a binary frame we will discuss the convexity of the belief space and some of its regions associated with important classes of b.f., namely Bayesian and simple support belief functions. Following this line of research in Section V we will prove that $\mathcal{B}$ has the form of a *polytope* or *simplex*, in which the basic probability assignment of a b.f. b plays the role of the simplicial coordinate of b in $\mathcal{B}$.

In Section VI we will instead focus our attention on the second fundamental notion of the ToE, Dempster's rule of combination. We will prove a fundamental result on Dempster's sums of convex combinations, and use it as a tool 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*.
Finally (Section VII) we will hint to some of the manifold open topics in the geometric approach, paying particular attention to the probabilistic approximation problem but without forgetting the geometry of possibility measures and the point-wise geometry of Dempster's rule.

II. The theory of evidence

Let us first review the basic notions of the theory of evidence.

A. Belief functions and basic probability assignments

Definition 1: A *basic probability assignment* (b.p.a.) over a finite set (*frame of discernment* [42]) Θ is a function $m : 2^\Theta \to [0,1]$ on its power set $2^\Theta \doteq \{A \subset \Theta\}$ such that

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

Subsets of Θ associated with non-zero values of m are called *focal elements*, and their union $\mathcal{C}$ *core*.

Definition 2: The *belief function* (b.f.) $b : 2^\Theta \to [0,1]$ associated with a basic probability assignment m_b on Θ is defined as:

$$b(A) = \sum_{B \subset A} m_b(B).$$

b can be viewed as the total probability induced by the generalized mass assignment m_b, even though an alternative definition can be given independently from the notion of basic probability assignment (see [42]).

In the simplest situation the evidence represented by a b.f. points to a single non-empty subset A of Θ.

Definition 3: A belief function $b : 2^\Theta \to [0,1]$ is called *simple support function* focused on A when $m_b(A) = \sigma$, $m_b(\Theta) = 1 - \sigma$, while $m_b(B) = 0$ for every other $B \subset \Theta$.
In the general case, though, a belief function supports more than one proposition. In particular, in the theory of evidence a finite probability function on Θ is just a peculiar b.f. (*Bayesian belief function*) assigning non-zero mass to elements of the frame only:

$$m_b(A) = 0 \quad \forall A : |A| > 1$$

(where $|A|$ denotes the cardinality of the subset A).

Belief functions admit the following order relation

$$b \leq b' \equiv b(A) \leq b'(A) \qquad \forall A \subset \Theta \tag{2}$$

called *weak inclusion*. A probability distribution p such that a belief function b is weakly included in p ($p(A) \geq b(A)\ \forall A$) is said to be *consistent* with b [30]: b can be viewed as the lower envelope of the set of probabilities consistent with it.

B. Dempster's rule

Belief functions representing distinct bodies of evidence can be combined by means of *Dempster's rule of combination*.

Definition 4: The *orthogonal sum* or *Dempster's sum* of two belief functions b_1, b_2 is a new belief function $b_1 \oplus b_2$ whose focal elements are all the possible non-empty intersections $A_i \cap B_j$ of focal elements of b_1 and b_2 respectively, and whose b.p.a. is given by

$$m_{b_1 \oplus b_2}(A) = \frac{\sum\limits_{i,j: A_i \cap B_j = A} m_{b_1}(A_i) m_{b_2}(B_j)}{\sum\limits_{i,j: A_i \cap B_j \neq \emptyset} m_{b_1}(A_i) m_{b_2}(B_j)} \tag{3}$$

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

We denote with $k(b_1, b_2)$ the denominator of Equation (3). The quantity $Con(b_1, b_2) \doteq 1 - k(b_1, b_2)$ measures the *degree of conflict* [42] of the two belief functions, i.e. the amount of probability they attribute to contradictory (i.e. disjoint) subsets. When $Con(b_1, b_2) = 1$ ($k(b_1, b_2) = 0$) the two functions cannot be combined.

Dempster's rule can be naturally extended to the combination of several belief functions.

III. A GEOMETRIC APPROACH TO THE THEORY OF EVIDENCE

When one tries and apply the theory of evidence to classical vision problems some important questions arise, stimulating major advances in the theory itself. Object tracking [35], for instance is one of the most active fields of computer vision, and concerns the reconstruction of the actual configuration or "pose" of a moving object by processing the sequence of images taken during its motion.

A. Object tracking and the probabilistic approximation problem

Let us assume that the pose of the moving object can be expressed as a point q of some region $\mathcal{Q}$ of $\mathbb{R}^D$, called *configuration space*. In case of rigid bodies the 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 also describes its internal configuration. Most of the literature concerns *model-based* approaches [25], in which some a-priori model of the body is used to help the estimation. In case of articulated objects composed by a number of rigid parts, a popular choice is to use a kinematic model.

However, when we have no a-priori information about the nature itself of the body (i.e. rigid, articulated, deformable, etc.), the only way of doing inference on the object pose is building a map between some salient image measurements (*features*) and poses in a learning stage.
In a training session the object describes a trajectory which approximates the actual parameter space $\mathcal{Q}$: a finite approximation $\tilde{\mathcal{Q}} = \{q_k, k = 1, ..., T\}$ of the parameter space is hence acquired as the collection of poses assumed by the object in the training session. These ground truth configuration values q_k can be produced for instance by a motion capture system, i.e. a machinery able to provide the 3D positions of reflective markers given the images coming from a set of calibrated cameras. At the same time, a number of features y_i are extracted from the available image sequences. For instance, in [15] we adopted color segmentation to separate the moving body (Figure 1-left) from a non-static background through a colorimetric analysis of the body of interest (Figure 1-right). We then found the bounding box (in red) around the the silhouette of the segmented upper torso or legs, and chose as features the coordinates of the vertices of the box. To

Fig. 1. Left. Training image of a person moving his right arm. Right. Feature extraction process: the object of interest is color segmented, and the bounding box (in red) around it is detected.

learn maps between features and poses, feature ranges are discretized by feeding a *hidden Markov model* (HMM) with the acquired measurements, yielding a partition $\{\mathcal{Y}^j, j = 1, ..., N\}$ (where N is the number of states of the HMM) of their ranges (see Figure 2-left). As a consequence,

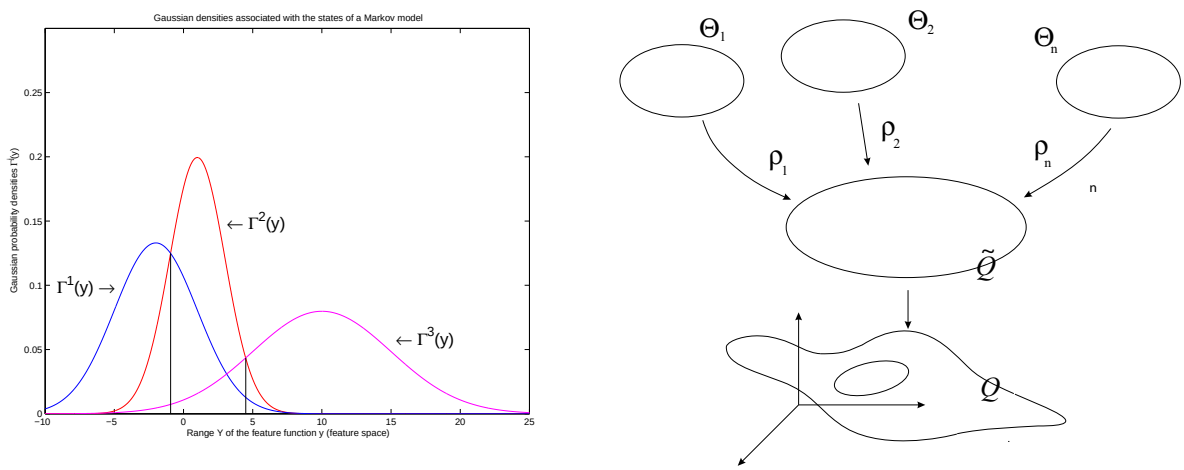

Fig. 2. Left. Implicit partition associated with a Markov model with three states. To each state corresponds a Gaussian density on the range $\mathcal{Y}$ of the feature function (abscissa). They define a partition of the range into three regions $\mathcal{Y}^1, \mathcal{Y}^2, \mathcal{Y}^3$ (separated by the vertical lines). Right. Evidential model architecture.

each element $\mathcal{Y}^j$ of this partition is mapped to the set of training poses whose feature value fall in $\mathcal{Y}^j$:

$$\rho : \mathcal{Y}^j \mapsto \{q_k : y(q_k) \in \mathcal{Y}^j\}.$$

In the evidential language, the maps ρ_i for all features i are *refinings*.

Definition 5: Given two frames Θ and Ω, a map $\rho : 2^\Theta \to 2^\Omega$ is a *refining* if it satisfies the following conditions: 1. $\rho(\{\theta\}) \neq \emptyset \ \forall \theta \in \Theta$; 2. $\rho(\{\theta\}) \cap \rho(\{\theta'\}) = \emptyset$ if $\theta \neq \theta'$; 3. $\cup_{\theta \in \Theta} \rho(\{\theta\}) = \Omega$.

The finer frame is called a *refinement* of the first one and Θ a *coarsening* of Ω.

The refinings ρ_i, together with discrete feature and parameter spaces, form the learned *evidential model* of the object (Figure 2-right). When the object evolves freely, the evidential model can be used to estimate its pose by encoding new features as belief functions, projecting them onto $\tilde{\mathcal{Q}}$, and combining them through Dempster's rule. This yields a belief estimate $\hat{b} : 2^{\tilde{\mathcal{Q}}} \to [0, 1]$ of the pose, which then needs to processed to extract a pointwise estimate $\hat{q}(t)$ of the configuration. A natural way is to approximate $\hat{b}$ with a finite probability $\hat{p}$ on $\tilde{\mathcal{Q}}$, and then compute its mean

value as

$$\hat{q} = \sum_{k=1}^{T} \hat{p}(q_k) \cdot q_k.$$

B. A geometric approach

An evidential solution to the object tracking problem then involves facing the probabilistic approximation problem. As we mentioned in the Introduction, the problem has been studied by many people. However, this can be posed in a different setting by investigating the shape of the space belief functions live, asking ourselves where do probability functions live in this space, and which is the correct distance to use to evaluate the difference between a belief function and a probability. The aim of this paper is to use the language of convex geometry to define a framework in which to answer all those questions.

The first pillar of the theory of evidence is the notion of basic probability assignment, i.e. the idea of assigning masses directly to events instead of elements of a frame. In the following two sections we are going to see that b.p.a.s translate naturally into the convex geometric language. We will first define the notion of *belief space* $\mathcal{B}$ as the space of all the belief functions on a given frame, and analyze the simplest case of a binary domain to get some intuitions about its properties for general frames. In particular we will observe that all probabilities live in a region which dominates the belief space (in the sense of Section II-A). The latter turns out to be convex, in analogy to a similar result for lower previsions [50].
In Section V instead, 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*.

Belief functions are useful when combined in an evidence revision process. The mechanism which regulates this process in the theory of evidence is Dempster's rule. In the second part of the paper we will then study the behavior of the rule of combination in our geometric framework, and give a description of the notion of conditional b.f. in geometric terms.

IV. THE 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 non-empty subsets of Θ, a reference frame (a set of linearly independent vectors)

$\{X_A : A \subset \Theta, A \neq \emptyset\}$. Each vector v of $\mathbb{R}^{N-1}$ can then be expressed in terms of this base as

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

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 belief function 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 $b(\emptyset) = 0$) then v is *potentially* a belief function, its component v_A measuring the belief value of A:

$$v_A = b(A) \quad \forall A \subset \Theta. \tag{4}$$

However, not every vector $v \in \mathbb{R}^{N-1}$ represents a valid belief function, as it may not meet the conditions of Definition 1.

Definition 6: The *belief space* associated with a frame Θ is the set of points $\mathcal{B}$ of $\mathbb{R}^{N-1}$ which correspond to a valid belief function (Equation (4)).

Let us study some of the properties of $\mathcal{B}$ to get first a rough idea of the shape of the belief space.

A. Belief space for a binary frame

To get some insight about the properties and geometric shape of the belief space it may be useful to have first a look at how belief functions 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 belief function $b : 2^{\Theta_2} \to [0, 1]$ 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 three-dimensional 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 $b(\Theta) = \sum_{A \subset \Theta} m_b(A) = 1$, the last coordinate of the vector can also be neglected (this is of course true for arbitrary frames too). In the binary case this means that we can represent b as the vector

$$[b(x) = m_b(x), b(y) = m_b(y)]' \in \mathbb{R}^2 \tag{5}$$

of $\mathbb{R}^{N-2} = \mathbb{R}^2$ (since $N = 2^2 = 4$).

Since $m_b(x) \geq 0$, $m_b(y) \geq 0$, and $m_b(x) + m_b(y) \leq 1$ we can easily infer that the set $\mathcal{B}_2$ of all the possible belief functions on Θ_2 can be depicted as the triangle in the Cartesian plane of Figure 3, whose vertices are the points

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

which correspond (through Equation (5)) respectively to the vacuous belief function 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 with $m_{b_y}(y) = 1$. The versors

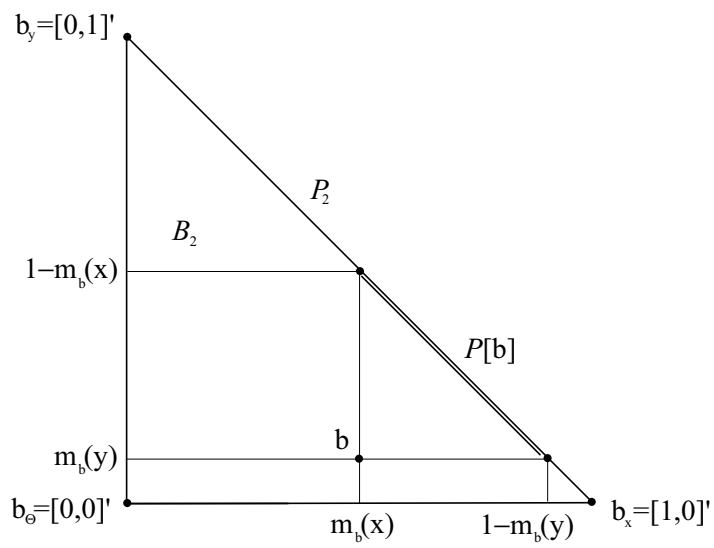

Fig. 3. The belief space $\mathcal{B}$ for a binary frame is a triangle in $\mathbb{R}^2$ whose vertices are the basis belief functions focused on $\{x\}, \{y\}$ and Θ, b_x, b_y, b_Θ respectively. The probability region is the segment $Cl(b_x, b_y)$, while the set of probabilities consistent with a b.f. b is also a segment, whose extreme points are given by Equation (7).

$X_x = [1,0]', X_y = [0,1]'$ form a reference frame $\{X_A : A \subset \Theta, A \neq \Theta, \emptyset\}$ in the Cartesian plane.

The Bayesian belief functions on Θ_2 obey the constraint

$$m_b(x) + m_b(y) = 1$$

and can then be located as points of the segment $\mathcal{P}_2$ joining $b_x = [1,0]'$ and $b_y = [0,1]'$.

Note that

$$\sum_{A \subset \Theta_2} b(A) = b(x) + b(y) + b(\Theta_2) = m_b(x) + m_b(y) + 1 = 2 - m_b(\Theta) \tag{6}$$

which is equal to 2 iff b is Bayesian.

The set of Bayesian b.f. consistent with b is the segment $P[b]$ in Figure 3 whose extreme points are the probabilities

$$[m_b(x), 1 - m_b(x)]', \qquad [1 - m_b(y), m_b(y)]'. \tag{7}$$

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 $m_b(\Theta)$ assigned to the whole frame.

From this example we can observe that

- 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 $\mathcal{B}, \mathcal{P}$;
- moreover, $\mathcal{B}, \mathcal{P}$ are both *polytopes* or *simplices*, i.e. convex closures of a finite sets of points

  $$\mathcal{B}_2 = Cl(b_\Theta, b_x, b_y) \qquad \mathcal{P}_2 = Cl(b_x, b_y)$$

  where the *convex closure* of a finite number of vectors $v_1, ..., v_k$ in a Cartesian space $\mathbb{R}^m$ is defined as

  $$Cl(v_1, ..., v_k) \doteq \Big\{ \sum_i \alpha_i v_i, \ \sum_i \alpha_i = 1, \alpha_i \geq 0 \Big\}. \tag{8}$$

- the set of probabilities consistent with a belief function b is also a simplex (a segment in the binary case).

We are going to see that these are general properties, valid for arbitrary frames.

B. The region of dominating probabilities

Let us start with a first characterization of the geometry of Bayesian belief functions. Let us denote with $n \doteq |\Theta|$ the cardinality of Θ. From the binary example we learned that that the probabilities on a frame Θ live in a well determined polytope in $\mathbb{R}^2$.

Generalizing condition (6) we can define

$$\mathcal{P} = \Big\{ b : 2^\Theta \to [0, 1] : \sum_{A \subset \Theta} b(A) = 2^{n-1} \Big\}. \tag{9}$$

and prove that

Theorem 1: The region of the belief space associated with all the Bayesian belief functions on Θ is the set (9).

Proof: If $b : 2^\Theta \to [0,1]$ is a belief function on a frame Θ

$$\sum_{A\subset\Theta} b(A) = \sum_{A\subset\Theta}\sum_{B\subset A} m_b(B) = \sum_{B\subset\Theta} m_b(B)|\{A : B \subset A \subset \Theta\}|$$

as each subset B is counted as many times as there are A's containing it. But since

$$|\{A : B \subset A \subset \Theta\}| = |\{C \subset (\Theta \setminus B)\}| = 2^{|\Theta\setminus B|} = 2^{n-|B|}$$

we have that (after switching back to the notation A for subsets of Θ)

$$\sum_{A\subset\Theta} b(A) = \sum_{A\subset\Theta} m_b(A)2^{n-|A|}. \tag{10}$$

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

Conversely, if

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

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

As the example of Section IV-A suggests, the belief space $\mathcal{B}$ is always dominated by the probability region $\mathcal{P}$.

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

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

where the equality holds iff b is Bayesian.

Proof: Recalling Equation (10), and after noticing that $2^{|\Theta\setminus B|} \leq 2^{n-1}$ (where the equality holds iff $|B| = 1$) we have that

$$\sum_{A\subset\Theta} b(A) = \sum_{A\subset\Theta} m_b(A)2^{|\Theta\setminus A|} \leq 2^{n-1}\sum_{A\subset\Theta} m_b(A) = 2^{n-1}\cdot 1$$

where the equality holds iff $|A| = 1$ for every focal element of b, i.e. b is Bayesian. ■

C. Convexity

The binary example also suggests that the belief space could be convex in the general case too. A similar result for a wider class of fuzzy measures, called coherent lower previsions, is in fact well known.

1) Convexity results for coherent lower previsions: It is well-known that belief functions are a special type of *coherent lower probabilities*, which in turn can be seen as a particular class of *lower previsions* (consult [50], section 5.13). More precisely,

Definition 7: Given a convex set of probability distributions K, then the function

$$\underline{P} = \inf_{p \in K} p(x)$$

is called *lower envelope* of K, or *coherent lower probability*.

A belief function is by definition the lower envelope of the set of probabilities consistent with it (2).

Definition 8: Suppose $\mathcal{A}$ is a collection of subsets of Θ and $\underline{P}$ a lower probability defined on $\mathcal{A}$ such that $\underline{P}(\emptyset) = 0$, $\underline{P}(\Theta) = 1$ and $\underline{P}(A) \geq 0$ for all $A \in \mathcal{A}$. Then $\underline{P}$ is called *completely monotone* if, for any collection $A_1, ..., A_k$ of elements of $\mathcal{A}$

$$\underline{P}(\bigcup_{i=1}^{k} A_i) \geq \sum_{J \subset \{1,2,...,k\}} (-1)^{|J|+1} \underline{P}(\bigcap_{j \in J} A_j).$$

The above condition clearly resounds the *superadditivity* axiom [42] of Shafer's definition of belief functions: for every positive integer n and every collection $A_1, ..., A_n \in 2^{\Theta}$

$$b(A_1 \cup ... \cup A_n) \geq \sum_i b(A_i) - \sum_{i<j} b(A_i \cap A_j) + ... + (-1)^{n+1} b(A_1 \cap ... \cap A_n). \quad (11)$$

Hence

Proposition 1: Belief functions are completely monotone coherent lower probabilities.

A proof of this result can be found in [50]. Walley has also proved that

Proposition 2: Coherent lower probabilities are closed under convex combination.

This implies that convex combinations of belief functions are still coherent. Here we are going to prove a stronger result.

2) Möbius inversion lemma and proof of convexity: Given a belief function b the corresponding basic probability assignment can be obtained by applying the *Möbius inversion lemma*[1]

$$m_b(A) = \sum_{B \subset A} (-1)^{|A \setminus B|} b(B). \quad (12)$$

[1]See [1] for an explanation in terms of the theory of monotone functions over partially ordered sets.

We can hence decide whether a point $v \in \mathbb{R}^{N-1}$ is a belief function by computing its b.p.a. and checking the axioms m_b must obey (see Definition 1).

The *normalization* constraint $\sum_{A \subset \Theta} m_b(A) = 1$ trivially translates into

$$\mathcal{B} \subset \{v \in \mathbb{R}^{N-1} : v_\Theta = 1\}. \tag{13}$$

As we have seen in the binary case, this implies that belief functions can indeed be represented as points of $\mathbb{R}^{N-2}$. From now on we will then think of $\mathcal{B}$ as a subset of $\mathbb{R}^{N-2}$.

The *positivity* condition instead

$$m_b(A) \geq 0 \qquad \forall A \subset \Theta,$$

originates the following inequality, which again recalls Equation (11)

$$\begin{aligned} m_b(A) = b(A) - \sum_{B \subset A, |B| = |A|-1} b(B) + \cdots + \\ +(-1)^{|A|-k} \sum_{|B|=k} b(B) + \cdots + (-1)^{|A|-1} \sum_{x \in \Theta} b(x) \geq 0 \qquad \forall A \subset \Theta. \end{aligned} \tag{14}$$

Now, all the constraints in Equation (14) have the form

$$\sum_{B \subset A, |A \setminus B| \; even} b(B) \geq \sum_{B \subset A, |A \setminus B| \; odd} b(B);$$

we can use them to prove that

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 $b_0, b_1 \in \mathcal{B}$

$$\sum_{B \subset A, |A \setminus B| \; even} b_0(B) \geq \sum_{B \subset A, |A \setminus B| \; odd} b_0(B), \quad \sum_{B \subset A, |A \setminus B| \; even} b_1(B) \geq \sum_{B \subset A, |A \setminus B| \; odd} b_1(B) \tag{15}$$

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

$$\begin{aligned} \sum_{B \subset A, |A \setminus B| \; even} b_\alpha(B) = \sum_{B \subset A, |A \setminus B| \; even} [b_0(B) + \alpha(b_1(B) - b_0(B))] = (1-\alpha) \sum_{B \subset A, |A \setminus B| \; even} b_0(B) + \\ +\alpha \sum_{B \subset A, |A \setminus B| \; even} b_1(B) \geq \end{aligned}$$

by Equation (15)

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

i.e. $b_\alpha \in \mathcal{B}$. ■

Theorem 3 is a strengthening of Proposition 2, stating that convex combinations of belief functions are not only coherent, but also completely monotone.

V. Simplicial form of the belief space

It is quite well-known that the set of probability distributions over a finite sample space Θ of cardinality n can be represented as 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,

$$[1, 0, ..., 0]', [0, 1, ..., 0]', \cdots, [0, ..., 0, 1]'.$$

On the other side, we have seen in Section IV-C that the belief space is convex. In fact, the binary example of Section IV-A has shown that $\mathcal{B}$ is a triangle (2-dim simplex) in the case of size 2 frames. We are ready to prove that the belief space can be given a polytope-like description for arbitrary frames too, generalizing the case of probability distributions [26].

A. Basis belief functions and simplex of belief functions

We have seen that in the binary example the belief space was a triangle, whose vertices were the following $N - 1 = 2^2 - 1 = 3$ vectors of $\mathbb{R}^2$

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

To prove that $\mathcal{B}$ is in fact a polytope in the general case too we first need to understand the geometric behavior of the basic probability assignment.

Theorem 4: The set of all the belief functions with focal elements in a given collection $\{A_1, ..., A_m\}$ is closed and convex in $\mathcal{B}$, namely

$$\{b : \mathcal{E}_b \subset \{A_1, ..., A_m\}\} = Cl(b_{A_i}, i = 1, ..., m)$$

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

$$b_A(B) = \begin{cases} 1 & if\ B \supset A \\ 0 & if\ B \not\supset A. \end{cases} \tag{16}$$

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

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

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

Each component $b(A)$ of these vectors b can be obviously written as

$$b(A) = \sum_{B \subset A, B \in \mathcal{E}_b} m_b(B) = \sum_{B \in \mathcal{E}_b} m_b(B) b_B(A)$$

where the "indicator" function

$$b_B(A) = \begin{cases} 1 & if\ A \supset B \\ 0 & if\ A \not\supset B. \end{cases}$$

selects the subsets B of A.

After collecting the values $b_B(A)$ in a vector $b_B = [b_B(A), A \subset \Theta, A \neq \emptyset, \Theta]'$ we can express b as a convex combination of the vectors b_B

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

as $m_b(B) = 0$ whenever $A_i \notin \mathcal{E}_b$. Since m_b is a basic probability assignment,

$$\textstyle\sum_{i=1}^{m} m_b(A_i) = 1, \qquad m_b(A_i) \geq 0\ \forall i$$

so that by definition of convex closure (8)

$$\Big\{ b : \mathcal{E}_b \subset \{A_1, ..., A_m\} \Big\} = \Big\{ b = \sum_{i=1}^{m} m_b(A_i) b_{A_i} \Big\} = Cl(b_{A_i}, i = 1, ..., m).$$

■

Proposition 3: The vector b_A defined by Equation (16) is the simple support b.f. assigning unitary mass to a single subset A:

$$m_{b_A}(A) = 1, \quad m_{b_A}(B) = 0 \quad \forall B \neq A; \tag{17}$$

we call it A-th *basis belief function*.

Proof: The belief values associated with the b.p.a. (17) are

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

which is exactly Equation (16). ■

Immediately, since $\mathcal{B}$ is the collection of belief functions b with focal elements in $2^\Theta \setminus \emptyset$ ($\mathcal{E}_b \subset 2^\Theta \setminus \emptyset$),

Corollary 1: The belief space $\mathcal{B}$ is the convex closure of all the basis belief functions,

$$\mathcal{B} = Cl(b_A, A \subset \Theta, A \neq \emptyset). \tag{18}$$

Note that even though the vectors $\{b_A\}$ are $N-2$-dimensional, $\mathcal{B}$ has $N-1$ vertices, including the basis b.f. b_Θ. Since $b_\Theta(B) = 0\ \forall B \subset \Theta, B \neq \Theta$ the vector $b_\Theta = \mathbf{0}$ coincides with the origin

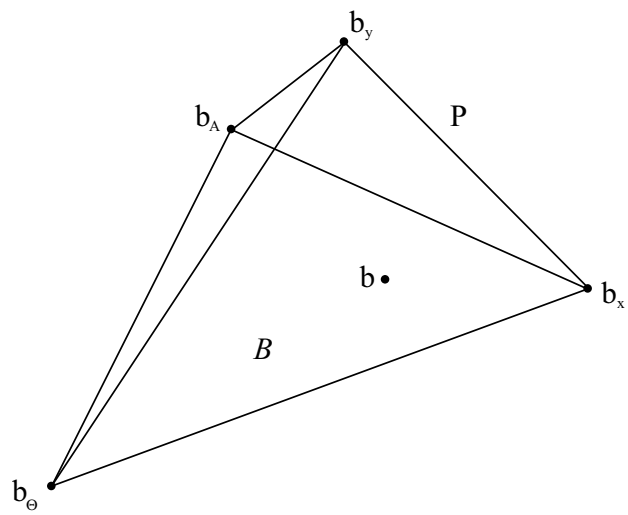

Fig. 4. Simplicial structure of the belief space $\mathcal{B} = Cl(b_A, A \neq \emptyset)$: its vertices are all the basis belief functions b_A represented as vectors of $\mathbb{R}^{N-2}$. The probabilistic subspace is a subset $\mathcal{P} = Cl(b_x, x \in \Theta)$ of its border.

of $\mathbb{R}^{N-2}$. The convex closure of a set of points is also called a *simplex*. Corollary 1 then states that $\mathcal{B}$ is the simplex whose vertices are all the basis b.f. The situation is illustrated in Figure 4. By Theorem 4 each belief function $b \in \mathcal{B}$ can be written as

$$b = \sum_{A \subset \Theta, A \neq \emptyset} m_b(A) b_A.$$

In other words, a basic probability assignment 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 belief function (a finite probability) is a b.f. with focal elements in the collection of the singletons: $\mathcal{C}_b = \{\{x_1\}, ..., \{x_n\}\}$. Immediately by Theorem 4

Corollary 2: The region of the belief space corresponding to probability functions is the part of its border determined by all the simple probabilities, i.e. the simplex[2]

$$\mathcal{P} = Cl(b_x, x \in \Theta).$$

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

The polytopes whose vertices are a subset of the collection of vertices of a simplex are called *faces* of the simplex. $\mathcal{P}$ is then a $(n-1)$-dimensional face of $\mathcal{B}$ (whose dimension is instead $N-2=2^n-2$ as it has 2^n-1 vertices).

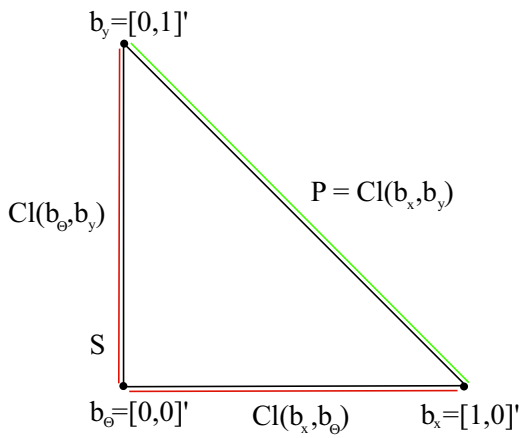

Fig. 5. Locations of the major classes of belief functions in the binary belief space.

On the other side, some one-dimensional faces of the belief space have also an intuitive meaning in terms of belief. Consider the segments $Cl(b_\Theta, b_A)$ joining the vacuous belief function b_Θ ($m_{b_\Theta}(\Theta)=1$,$m_{b_\Theta}(B)=0\ \forall B\neq\Theta$) with the basis b.f. b_A (16). 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 belief function b has b.p.a.

$$m_b(A)=\alpha, \qquad m_b(\Theta)=1-\alpha$$

i.e. b is a simple support function focused on A (Definition 3).
Accordingly, the union of these segments for all events A

$$\mathcal{S}=\bigcup_{A\subset\Theta} Cl(b_\Theta, b_A)$$

is the region of simple support belief functions on Θ.

Let us consider again the binary case (Figure 5). 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 one-dimensional polytope. As a matter of fact, Ha *et al.* [26] proved that $P[b]$ can be expressed in the probability simplex as the sum of the probability simplices associated to the focal elements A_i, $i = 1, \cdots, k$ of b, weighted by the corresponding masses, i.e.

$$P[b] = \sum_{i=1}^{k} m_b(A_i) conv(A_i) \tag{19}$$

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. More formally, to each focal element A corresponds a mass $m_b(A)$ distributed among its elements, $m_b(A) \cdot Cl(b_x, x \in A)$, so that $P[b]$ can be expressed as in (19) in an arbitrary belief space.

Let us then find an explicit expression for (19). Given an arbitrary belief function b with focal elements $A_1, ..., A_k$, we can define for each choice of k representatives $\{x_1, ..., x_k\}$, $x_i \in A_i \ \forall i$ of the focal elements, the b.f.

$$b(x_1, ..., x_k) \doteq \sum_{i=1}^{k} m_b(A_i) b_{x_i}. \tag{20}$$

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

Proof: Starting from Equation (19), $P[b]$ can be developed as

$$\begin{aligned}
P[b] &= \Big\{ \sum_{i=1}^{k} m_b(A_i) \Big(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \Big), \sum_{j=1}^{|A_i|} \alpha_i^j = 1 \ \forall i \Big\} = \\
&= \Big\{ m_b(A_1) \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} b_{x_1^{j_1}} + \sum_{i=2}^{k} m_b(A_i) \Big(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \Big), \sum_{j=1}^{|A_i|} \alpha_i^j = 1 \ \forall i \Big\} = \\
&= \Big\{ m_b(A_1) \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} b_{x_1^{j_1}} + \sum_{i=2}^{k} \Big(\sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} \Big) m_b(A_i) \Big(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \Big), \sum_{j=1}^{|A_i|} \alpha_i^j = 1 \ \forall i \Big\} = \\
&= \Big\{ \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} m_b(A_1) b_{x_1^{j_1}} + \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} \Big(\sum_{i=2}^{k} m_b(A_i) \Big(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \Big) \Big), \sum_{j=1}^{|A_i|} \alpha_i^j = 1 \ \forall i \Big\} = \\
&= \Big\{ \sum_{j_1=1}^{|A_1|} \alpha_1^{j_1} \Big[m_b(A_1) b_{x_1^{j_1}} + \sum_{i=2}^{k} m_b(A_i) \Big(\sum_{j=1}^{|A_i|} \alpha_i^j b_{x_i^j} \Big) \Big], \sum_{j=1}^{|A_i|} \alpha_i^j = 1 \ \forall i \Big\}
\end{aligned} \tag{21}$$

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

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

which replaced in (21) yields

$$P[b] = \Big\{\sum_{j_1=1}^{|A_1|}\sum_{j_2=1}^{|A_2|}\beta_{j_1j_2}\Big[m_b(A_1)b_{x_1^{j_1}} + m_b(A_2)b_{x_2^{j_2}} + \sum_{i=3}^{k} m_b(A_i)\Big(\sum_{j=1}^{|A_i|}\alpha_i^j b_{x_i^j}\Big)\Big] :$$
$$: \sum_{j_1j_2}\beta_{j_1j_2} = 1, \sum_{j=1}^{|A_i|}\alpha_i^j = 1\ \forall i = 3, ..., k\Big\}$$

with $\beta_{j_1j_2} \doteq \alpha_1^{j_1}\cdot\alpha_2^{j_2}$. Clearly the expression inside the square brackets has the same shape as above, 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

$$\frac{1}{\prod_i |A_i|}\sum_{\{x_1,...,x_k\}\in A_1\times...\times A_k} b_(x_1, ..., x_k) = \frac{1}{\prod_i |A_i|}\sum_{\{x_1,...,x_k\}\in A_1\times...\times A_k}\sum_{i=1}^{k} m_b(A_i)b_{x_i} =$$
$$= \frac{1}{\prod_i |A_i|}\sum_{a\in\mathcal{C}_s} b_a\cdot\sum_{A_j\supset a} m(A_j)\frac{\prod_i |A_i|}{|A_j|} \tag{22}$$

as each basis probability b_x appears in (22) with coefficient $m_b(A_j)$ a number of times $\prod_{i\neq j}|A_i|$ equal to the number of possible choices of the representatives for the other focal elements of b. This in turn reads as

$$\sum_{a\in\mathcal{C}_s} b_a\cdot\sum_{A_j\supset a}\frac{m(A_j)}{|A_j|} = \sum_{x\in\Theta} b_x\cdot\sum_{A\supset x}\frac{m(A)}{|A|} \tag{23}$$

(since no focal elements include points outside the core), which is nothing but the pignistic function.

The geometric analysis of the convex 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. 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 [40], 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 (22).

VI. Geometry of Dempster's rule

In the first part of the paper we investigated the geometric properties of the first pillar of the theory of evidence, the twin notions of belief function and basic probability assignment. We now know that belief functions live in a simplex in the Cartesian space $\mathbb{R}^{2^{|\Theta|}-2}$ whose vertices represent b.f. focused on a single event, and where b.p.a.s can be interpreted as simplicial coordinates of the points representing them.

It is now time to study the geometry of the other key concept of the theory of evidence, Dempster's rule of combination, which is in turn related to the notion of *conditional* belief function.

In the literature, in fact, conditional belief functions have been given several alternative definitions by different authors [41]. Fagin and Halpern, for instance, defined a notion of *conditional belief* [24] as the lower envelope of a family of conditional probability functions, and provided a closed-form expression for it. On the other side, M. Spies [47] 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 multi-valued mapping (which induces a belief function according to Dempster's original formulation) he gave a new definition of conditional belief function. An updating rule equivalent to the law of total probability when all beliefs are probabilities was introduced.

In [43] Slobodova described instead how conditional belief functions (defined as in Spies' approach) fit in the framework of valuation-based systems, while Xu and Smets ([51], [52]) showed how to use conditional belief functions to represent relations among variables as joint belief functions 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 [2], and the nature of belief propagations in evidential networks has been investigated [54].

In the following we will call conditional belief function $b|b'$ the combination of b with b'

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

In this form conditional belief functions arise from the application of the theory of evidence to estimation problems in which some sort of "temporal coherence" has to be enforced. *Data association* is a typical example.

A. Data association and conditional belief functions

In the *data association* problem a number of points moving in the 3D space are tracked by one or more cameras and appear in an image sequence as unlabeled (undistinguishable) points, and we seek for the correspondences between points of two consecutive frames. A popular approach called *joint probabilistic data association* filter [3] 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 small region the algorithm cannot distinguish between them.

However, when additional information is available it can be used to help the association process. One way to do this is representing the evidence combing from Kalman filters and other available constraints on the targets' motion as belief functions, and combining them on the space of all the 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 (see Figure 6-left)), then the rigid motion constraint can be exploited to improve the robustness of the estimation.

Formally, let us call the set of points of the known model $\{M_j, j = 1, \cdots, N\}$, and $\{m_l^k, l = 1, \cdots, n(k)\}$ the measured feature points in time k image (where $n(k)$ is the number of detected feature points). The data association problem the 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 safely assume that $n(k) = N$. The information carried by predictions of Kalman filters 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. on the frame of all *feature-to-feature* associations:

$$\Theta_k^{k-1} \doteq \{m_l^{k-1} \leftrightarrow m_m^k,\ \forall l, m = 1, \cdots, N\}$$

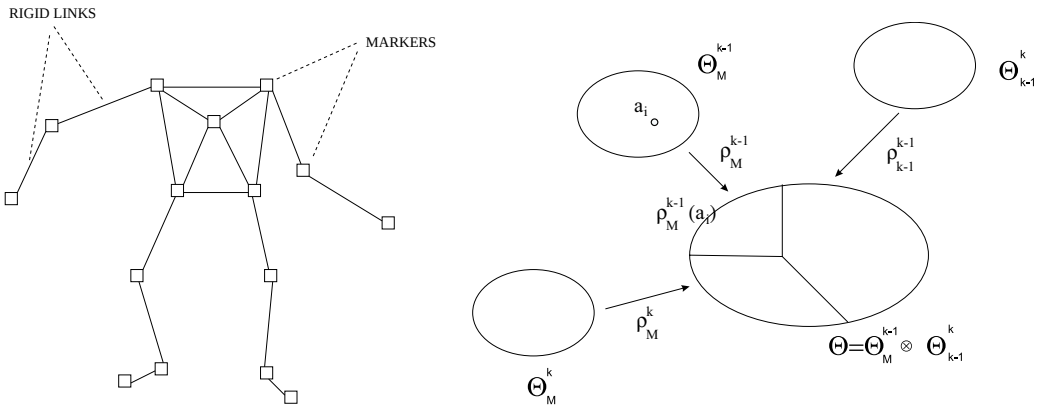

Fig. 6. (Left) Topological model of a human body: rigid motion constraints between pairs of markers are shown as links. (Right) The family of frames involved in the data association problem: all the constraints (expressed as b.f.) are combined over the common refinement Θ and then projected onto the current-time association frame Θ_M^k.

The rigid motion constraint depends on the other hand on the *model-measurement association* at the previous step: those associations are collected in the frame of all *past model-feature* associations,

$$\Theta_M^{k-1} \doteq \{m_l^{k-1} \leftrightarrow M_j,\ \forall j, l = 1, \cdots, N\}$$

while at each time instant k the desired associations $m_l^k \leftrightarrow M_j$ are elements of the *current model-feature* associations frame

$$\Theta_M^k \doteq \{m_l^k \leftrightarrow M_j,\ \forall j, l = 1, \cdots, N\}.$$

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

Definition 9: Two belief functions b_1, b_2 defined on two frames Θ and Ω connected by a refining ρ are said *consistent* iff

$$m_{b_1}(A) = m_{b_2}(\rho(A)), \qquad \forall A \subset \Theta$$

and b_1 is called the *restriction* of b_2 to Θ.

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 belief functions

$$b_i : 2^{\rho_M^{k-1}(\{a_i\})} \rightarrow [0,1]$$

where a_i is the i-th 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 its coarsening Θ_M^{k-1} (see Figure 6-right).

These conditional belief functions b_i must be reduced to a single *total belief function*, that will be eventually pooled with the other constraints, stimulating the search for the generalization of the total probability theorem to belief functions. 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 belief function defined on Ω and $\{b_1, ..., b_n\}$ a collection of n belief functions $b_i : 2^{\Theta_i} \rightarrow [0,1]$ defined on the elements Θ_i the partition $\{\Theta_1, ..., \Theta_n\}$ of Θ induced by the coarsening Ω.

Then there exists a belief function over $b : 2^\Theta \rightarrow [0,1]$ such that:

1) *a-priori constraint*: b_0 is the restriction [42] of b to Ω;
2) *conditional constraint*: the *conditional* belief function obtained by combining b with

$$b_{\Theta_i} : m_{b_{\Theta_i}}(\Theta_i) = 1, m_{b_{\Theta_i}}(A) = 0 \forall A \subset \Theta_i, A \neq \Theta_i$$

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

The hypotheses of Theorem 6 are pictorially summarized in Figure 7. In the data association problem the a-priori constraint is the belief function representing the estimate of the past association $\{m_l^{k-1} \leftrightarrow M_j\}$, defined over Θ_k^{k-1} (see Figure 6-right again), and ensures that the total belief function 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 instead use the language of convex geometry we introduced in the first part to give a characterization of the

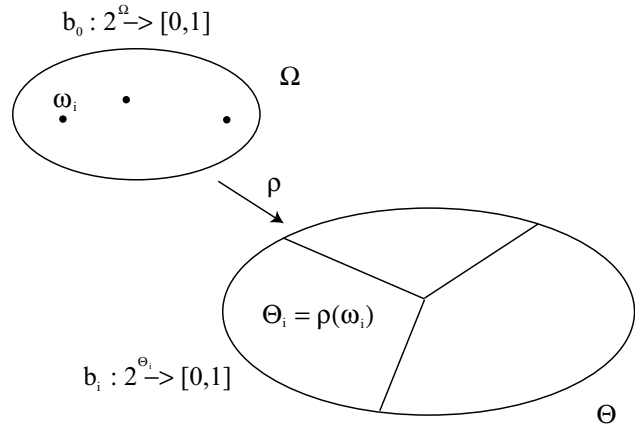

Fig. 7. The total belief theorem: a belief function b on Θ such that its restriction to Ω is b_0 and whose combination with b_{Θ_i} (where $\{\Theta_1, \cdots, \Theta_n\}$ is the partition of Θ induced by the refining ρ) is b_i is desired.

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 as a tool to show that the rule of combination commutes with the convex closure operator in the belief space. This will finally allow us to describe the "global" geometry of the orthogonal sum in terms of simplices called *conditional subspaces*, i.e. the sets of all belief functions conditioned by a given b.f. b.

Theorem 7: Consider a belief function b and a collection of b.f. $\{b_1, ..., b_n\}$ such that there exists at least a belief function b_j combinable with b. If $\sum_i \alpha_i = 1$, $\alpha_i > 0$ for all $i = 1, ..., n$ then

$$b \oplus \sum_i \alpha_i b_i = \sum_i \beta_i (b \oplus b_i)$$

with

$$\beta_i = \frac{\alpha_i k(b, b_i)}{\sum_{j=1}^n \alpha_j k(b, b_j)}, \tag{24}$$

where $k(b, b_i)$ is the normalization factor for the i-th combination $b \oplus b_i$:

$$k(b, b_i) \doteq \sum_{A \cap B \neq \emptyset} m_b(A) m_{b_i}(B).$$

Proof: We just need to check the equality of the corresponding basic probability assignments. After denoting with $\{B_k\}$ the focal elements of b_i and with $\{A_j\}$ those of b, we have

that 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

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

On the other side, by hypothesis

$$m_{\sum_i \alpha_i b_i}(A) = \sum_{B \subset A} (-1)^{|A-B|} \sum_i \alpha_i b_i(B) = \sum_i \alpha_i \Big(\sum_{B \subset A} (-1)^{|A-B|} b_i(B) \Big) = \sum_i \alpha_i m_i(A).$$

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

$$m_{b \oplus \sum_i \alpha_i b_i}(A) = \frac{\sum\limits_{E_k \cap A_j = A} m_{\sum_i \alpha_i b_i}(E_k) m_b(A_j)}{\sum\limits_{E_k \cap A_j \neq \emptyset} m_{\sum_i \alpha_i b_i}(E_k) m_b(A_j)} = \frac{\sum\limits_{E_k \cap A_j = A} \Big(\sum\limits_i \alpha_i m_{b_i}(E_k) \Big) \cdot m_b(A_j)}{\sum\limits_{E_k \cap A_j \neq \emptyset} \big(\sum\limits_i \alpha_i m_{b_i}(E_k) \big) \cdot m_b(A_j)} =$$

$$= \frac{\sum\limits_i \alpha_i \Big(\sum\limits_{E_k \cap A_j = A} m_{b_i}(E_k) m_b(A_j) \Big)}{\sum\limits_i \alpha_i \Big(\sum\limits_{E_k \cap A_j \neq \emptyset} m_{b_i}(E_k) m_b(A_j) \Big)} = \frac{\sum\limits_i \alpha_i \Big(\sum\limits_{B_k \cap A_j = A} m_{b_i}(B_k) m_b(A_j) \Big)}{\sum\limits_i \alpha_i \Big(\sum\limits_{B_k \cap A_j \neq \emptyset} m_{b_i}(B_k) m_b(A_j) \Big)} \tag{25}$$

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 Equation (25) 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 Equation (24). Now, if there is a b.f. b_j in the collection $\{b_1, ..., b_n\}$ which is combinable with b, then $k(b, b_j) \neq 0$ and the denominator of the above equation is non-zero, i.e. $m_{b \oplus \sum_i \alpha_i b_i}$ is well defined. ■

Note that since

$$\sum_i \beta_i = 1, \qquad \beta_i \geq 0 \ \ \forall i$$

the combination of b with any convex sum of belief functions is still a convex sum of all the partial combinations.

As an example, let us consider three belief functions 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

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

and its combination with b yields

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

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 $\beta_1 = \frac{\alpha k(b,b_1)}{\alpha k(b,b_1)+(1-\alpha)k(b,b_2)}$, $\beta_2 = \frac{(1-\alpha)k(b,b_2)}{\alpha k(b,b_1)+(1-\alpha)k(b,b_2)}$, 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 belief functions as points of the belief space.
It is not surprising to see that 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. can be exchanged. We just need to pay some attention to the issue of combinability.

Theorem 8: $b \oplus Cl(b_1, ..., b_k) = Cl(b \oplus b_{i_1}, ..., b \oplus b_{i_m})$ where $\{b_{i_1}, ..., b_{i_m}\} \subset \{b_1, ..., b_k\}$ are all the belief functions in the collection $\{b_1, ..., b_k\}$ which are combinable with b.

Proof: Sufficiency. We need to prove that if $b' \in b \oplus Cl(b_1, ..., b_k)$ then $b' \in Cl(b \oplus b_{i_1}, ..., b \oplus b_{i_m})$. If $b' = b \oplus \sum_{i=1}^{k} \alpha_i b_i$, $\sum_i \alpha_i = 1$ then (by Theorem 7) $b' = \sum_i \beta_i b \oplus b_i$ with β_i given by Equation (24). But we know that $\beta_i = 0$ iff $\nexists b \oplus b_i$, so that

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

Necessity. We have to show that if $b' \in Cl(b \oplus b_{i_1}, ..., b \oplus b_{i_m})$ then $b' \in b \oplus Cl(b_{i_1}, ..., b_{i_m}, b_{j_1}, ..., b_{j_l})$ for each choice of $b_{j_1}, ..., b_{j_l}$ not combinable with b.

If $\sum_{p=1}^{m} \alpha_p + \sum_{q=1}^{l} \alpha_q = 1$

$$b' = b \oplus \Big(\sum_{p=1}^{m} \alpha_p b_{i_p} + \sum_{q=1}^{l} \alpha_q b_{j_q} \Big) = \sum_{p=1}^{m} \alpha'_p b \oplus b_{i_p} + \sum_{q=1}^{l} \alpha'_q b \oplus b_{j_q} \tag{26}$$

with

$$\alpha'_p = \frac{\alpha_p \Delta_p}{\sum_p \alpha_p \Delta_p + \sum_q \alpha_q \Delta_q}, \quad \alpha'_q = \frac{\alpha_q \Delta_q}{\sum_p \alpha_p \Delta_p + \sum_q \alpha_q \Delta_q}$$

from Equation (24). But now $\Delta_q = 0 \ \forall q$ (as b_{j_q} is not combinable with b) so that $\alpha'_q = 0$ for all $q = 1, ..., l$, and from Equation (26) it follows that

$$b' = \sum_{p=1}^{m} \beta_p \cdot b \oplus b_{i_p} \in Cl(b \oplus b_{i_1}, ..., b \oplus b_{i_m}), \tag{27}$$

with $\beta_p = \frac{\alpha_p \Delta_p}{\sum_p \alpha_p \Delta_p}$, $\sum_p \beta_p = 1$.

Hence, any belief function b' of the form (27) belongs to region $b \oplus Cl(b_{i_1}, ..., b_{i_m}, b_{j_1}, ..., b_{j_l})$ iff we can find another collection of coefficients $\{\alpha_p, \ p = 1, ..., m\}$ with $\sum_p \alpha_p = 1$ such that the following constraints are met

$$\beta_p = \frac{\alpha_p \Delta_p}{\sum_p \alpha_p \Delta_p} \quad \forall p = 1, ..., m \tag{28}$$

(i.e. $b' = b \oplus \sum_p \alpha_p b_{i_p}$).

An admissible solution to the system of equations (28) is $\tilde{\alpha}_p \doteq \beta_p / \Delta_p$ as we get $\forall p \ \beta_p = \beta_p / \sum_p \beta_p = \beta_p$ since the β_p's are normalized, and system (28) is satisfied up to the normalization constraint. We can then normalize the solution by choosing

$$\alpha_p = \tilde{\alpha_p} / \sum_{p'} \tilde{\alpha_{p'}} = \frac{\beta_p}{\Delta_p \sum_{p'} (\frac{\beta_{p'}}{\Delta_{p'}})}$$

so that system (28) is still met, together with the normalization condition. ■

An immediate consequence is that

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

D. Conditional subspaces

As basically a linear operator on $\mathcal{B}$, Dempster's rule commute with convex closure (Corollary 3). This is fundamental in the framework of the geometric approach, where all the major classes of b.f. form some sort of simplices. Furthermore, using the language the above commutativity

results provide we can also identity geometric counterparts of the notions of combinability and conditioning. The basic notion is that of conditional subspace.

Definition 10: The *conditional subspace* $\langle b \rangle$ associated with a belief function b is the set of all the belief functions *conditioned by* b, namely

$$\langle b \rangle \doteq \left\{ b \oplus b', \forall b' \in \mathcal{B} \; s.t. \; \exists \; b \oplus b' \right\}. \tag{29}$$

In a sense, the conditional subspace $\langle b \rangle$ is the possible "future" of b when combined by means of Dempster's rule. In a process of knowledge accumulation in which new evidence becomes available in the form of a belief function (and is combined through Dempster's rule)

$$b_{t_0}, b_{t_0} \oplus b_{t_1}, b_{t_0} \oplus b_{t_1} \oplus b_{t_2}, \cdots$$

the conditional subspace of the current knowledge state $\langle b_{t_0} \oplus \cdots \oplus b_t \rangle$ constrains the possible outcomes of the future states of belief.

However, since belief functions are not necessarily combinable, before we study $\langle b \rangle$ we need to understand the geometry of the notion of combinability.

Definition 11: The *non-combinable region* $NC(b)$ associated with a belief function b is the collection of all the b.f. which 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 geometric shape of this set. As a matter of fact the non-combinable region $NC(b)$ of b is also a simplex, whose vertices are the basis belief functions related to subsets disjoint from the core of b (the union of its focal elements).

Proposition 4: $NC(b) = Cl(b_A, A \cap \mathcal{C}_b = \emptyset)$.

Proof: It suffices to point out that

$$NC(b) = \{b' : \mathcal{C}_{b'} \subset \overline{\mathcal{C}}_b\} = \{b' : \mathcal{E}_{b'} \subset 2^{\bar{\mathcal{C}}_b}\}$$

where $\overline{B}$ denotes the complement of a subset B of Θ. But by Theorem 4

$$\{b' : \mathcal{E}_b \subset 2^{\bar{\mathcal{C}}_b}\} = Cl(b_A, A \in 2^{\bar{\mathcal{C}}_b}) = Cl(b_A : A \subset \bar{\mathcal{C}}_b)\} = Cl(b_A : A \cap \mathcal{C}_b = \emptyset).$$

■

Using the definition of non-combinable 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, i.e. the collection of belief functions in $\mathcal{B} \setminus NC(b)$ cannot be written as a set of b.f. 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 non-empty intersection with $\mathcal{C}_b$, regardless the behavior of the others.

Geometrically, as we know by Theorem 4, 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 12: The *compatible simplex* $C(b)$ associated with a belief function b is the collection of all the b.f. whose focal elements are in the core of b:

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

Now from Theorem 4 it follows that

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

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

Theorem 9:

$$\langle b \rangle = b \oplus C(b).$$

Proof: Let us denote with $\mathcal{E}_b = \{A_i\}$ and $\mathcal{E}_{b'} = \{B_j\}$ 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

$$\{B'_k, \ k = 1, ..., m\} \doteq \{B_j \cap \mathcal{C}_b, \ j = 1, ..., |\mathcal{E}_{b'}|\}$$

(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 $b \oplus b' = b \oplus b''$. ■

An analogous result can be found in [42].

We are now ready to understand the convex geometry of conditional subspaces. From Theorems 4 and 9 it follows that

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

Note that, since $b \oplus b_{\mathcal{C}_b} = b$ (where $b_{\mathcal{C}_b}$ is the basis belief function focused on the core of b), b is always one of the vertices of $\langle b \rangle$. Furthermore, $\langle b \rangle \subset C(b)$, since the core of a belief function b is such that [42] $\mathcal{C}_{b \oplus b'} = \mathcal{C}_b \cap \mathcal{C}_{b'} \subset \mathcal{C}_b$.

1) Example: binary frame: Figure 8 shows the actual shape of a conditional subspace for a b.f. defined on the simplest (binary) frame $\Theta_2 = \{x, y\}$. For each belief function $b \in \mathcal{B}_2$, $b \neq b_x, b_y$, the non-combinable 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

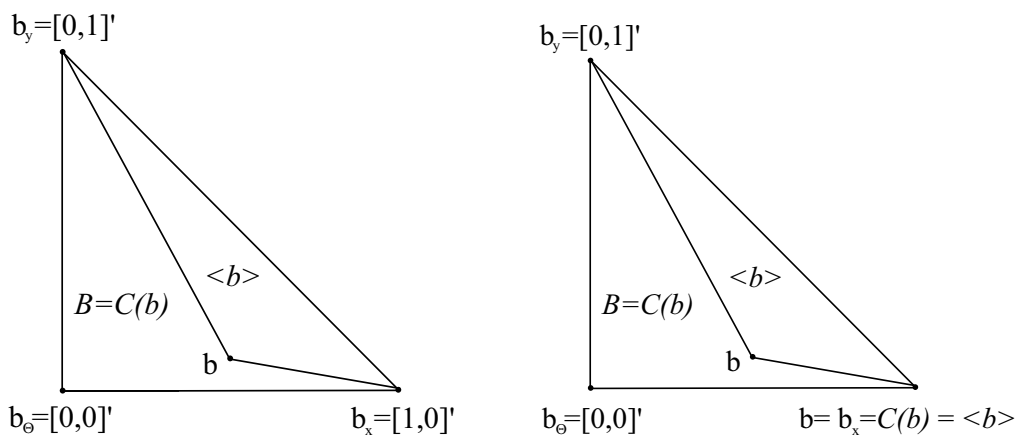

Fig. 8. The conditional subspace $\langle b \rangle$ of a belief function b in the binary belief space $\mathcal{B}_2$, along with its compatible subspace. On the abscissa we have the belief value $b(x)$ of x, while $b(y)$ is the coordinate of b on the y axis. $b(\Theta) = 1 = const$ is neglected. Left: 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.

the conditional subspace $\langle b \rangle$ are then

$$b \oplus b_\Theta = b, \quad b \oplus b_x = b_x, \quad b \oplus b_y = b_y$$

and $\langle b \rangle$ is the simplex depicted in Figure 8-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. 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 Figure 8-right).

VII. Applications and developments of the geometric approach

To conclude, we would like to give a flavor of a number of natural applications of the geometric approach to the theory of evidence we presented in this paper, together with a hint of the possible future developments of this framework.

A. Probabilistic approximation of a belief function

We already mentioned that one of the original motivations of this work was the probabilistic approximation problem, i.e. the problem of finding the probability which is the "closest" in some sense to a given belief function.

1) L_1 distance in the belief space: We can first consider the possibility of using the L_1 norm to measure distances in the belief space,

$$\|b - b'\|_{L_1} \doteq \sum_{A \subset \Theta} |b(A) - b'(A)|.$$

Unfortunately, it turns out to be of no use since, as Theorem 10 shows, all the Bayesian functions consistent with b have the same L_1 distance from b.

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

Proof: Obviously, since $b(A) \geq b'(A)$ for every $A \subset \Theta$, that is true for $\mathcal{C}_{b'} \subset \Theta$ 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 $A \supset \mathcal{C}$) this is equivalent to $\mathcal{C}_{b'} \supset \mathcal{C}_b$. ■

Theorem 10: If $b : 2^\Theta \to [0, 1]$ is an arbitrary belief function 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

$$\|p - b\|_{L_1} \doteq \sum_{A \subset \Theta} |p(A) - b(A)| = 2^{|\Theta \setminus \mathcal{C}_b|} \cdot \Big[2^{|\mathcal{C}_b|-1} - 1 - \sum_{A \subset \mathcal{C}_b, A \neq \mathcal{C}_b} b(A)\Big]. \tag{30}$$

Proof: Lemma 1 guarantees that $\mathcal{C}_p \subset \mathcal{C}_b$, so that $p(A) - b(A) = 1 - 1 = 0$ for $A \supset \mathcal{C}_b$. 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 which correspond to unions of non empty proper subsets of $\mathcal{C}_b$ and arbitrary subsets of $\bar{\mathcal{C}}_b = \Theta \setminus \mathcal{C}_b$. Notice that, by definition, if $A' = A \cup B$, with $A \subset \mathcal{C}_b, B \subset \bar{\mathcal{C}}_b$, we have $b(A') = b(A)$ (see Theorem 9).

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

$$\sum_{A \subset \Theta} |p(A) - b(A)| = \sum_{A' = A \cup B, A \subset \mathcal{C}_b, B \subset \bar{\mathcal{C}}_b} |p(A') - b(A')| = \sum_{A \subset \mathcal{C}_b} 2^{|\Theta \setminus \mathcal{C}_b|} |p(A) - b(A)| =$$
$$= 2^{|\Theta \setminus \mathcal{C}_b|} \sum_{A \subset \mathcal{C}_b} |p(A) - b(A)| = 2^{|\Theta \setminus \mathcal{C}_b|} \cdot \Big[\sum_{A \subset \mathcal{C}_b, A \neq \mathcal{C}_b} p(A) - \sum_{A \subset \mathcal{C}_b, A \neq \mathcal{C}_b} b(A)\Big]$$

since $p(A) \geq b(A)\ \forall A$. But then, by Equation (10)

$$\sum_{A \subset \mathcal{C}_b, A \neq \mathcal{C}_b} p(A) = 2^{|\mathcal{C}_b|-1} - p(\mathcal{C}_b) = 2^{|\mathcal{C}_b|-1} - 1$$

and we get Equation (30). ■

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

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

In [12] we studied the problem of finding the expression of $\pi[b]$ in terms of the belief values of the original belief function b, and proved a number of properties it possesses. In particular, we showed that the orthogonal projection commutes with convex combination,

$$\pi[\textstyle\sum_i \alpha_i b_i] = \sum_i \alpha_i \pi[b_i], \qquad \sum_i \alpha_i = 1$$

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 theory of evidence: a belief function is useful only when it is combined with others in a reasoning process. It is natural to think that this should be taken into account when tacking 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 belief function, when combined with any other b.f., must produce results similar to those obtained by combining the original belief function.*

Analytically, this yields the following optimization problem

$$\hat{b} = \arg\min_{b'' \in \mathcal{A}} \int_{b' \in \mathcal{B}} dist(b \oplus b', b'' \oplus b') db' \tag{31}$$

where b is the original belief function to approximate, $b' \in \mathcal{B}$ is an arbitrary belief function on the same frame, $dist$ is some distance function, and $\mathcal{A}$ is the class of belief functions the

approximation belongs to. Of course the role of $\oplus$ can be played by any other meaningful operator, like for instance the disjunctive rule of combination for unnormalized belief functions [45]. Possibly, the resulting approximation should be independent from the choice of this distance used in (31).

Let us consider here the class $\mathcal{A} = \mathcal{P}$ of the Bayesian belief functions. Since $\tilde{pl}_b$ represents perfectly b when combined with any Bayesian function [49]

$$b \oplus p = \tilde{pl}_b \oplus p \qquad \forall p \in \mathcal{P}$$

the modified version of the approximation problem (31) in which the original b.f. is combined with Bayesian belief functions only is trivially solved by $\tilde{pl}_b$:

$$\tilde{pl}_b = \arg\min_{p \in \mathcal{P}} \int_{p' \in \mathcal{P}} \|b \oplus p' - p \oplus p'\| dp' \tag{32}$$

whatever the norm we choose, as $b \oplus p' - p \oplus p' = 0 \ \forall p'$. It is then natural to conjecture that the relative plausibility function could be the solution of the general approximation problem (31), too. We will work on this conjecture in the near future.

B. The geometry of possibility measures

A consonant belief function is a b.f. whose focal elements are nested. All the possible lists of focal elements associated with consonant belief functions then correspond to all the possible chains of subsets

$$A_1 \subset ... \subset A_m$$

of Θ. Theorem 4 then implies that all the b.f.s whose focal elements belong to a chain $X = \{A_1, ..., A_m\}$ is $Cl(b_{A_1}, ..., b_{A_m})$. The region of the belief space formed by consonant belief functions turns then out to be the union of a collection of convex components, each associated with a *maximal* chain $\mathcal{A}$:

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

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 $\dim Cl(b_{A_1}, ..., b_{A_n}) = n - 1$.

In [14] 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; 2. the intersection of two d-dimensional simplices is a face of both the intersecting simplices.

The geometric description of consonant belief functions in the belief space 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 connection of those measures with 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 [23], [4]) of a belief function b, according to the optimization criterion (31), the unique consonant belief function c with plausibility

$$pl_c(A) = \frac{\max_{x \in A} pl_b(x)}{\max_{x \in \Theta} pl_b(x)}. \tag{33}$$

In the near future we are going to work on a formal proof of this conjecture.

C. Canonical decomposition

A large class of belief functions is composed by all the b.f. which are the result of a Dempster's sum of simple support b.f.

Definition 13: A *separable support function* is a belief function which 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. $\forall i = 1, .., n$.

Separable support functions can be decomposed in different ways. However [42],

Proposition 5: If $b \neq b_\Theta$ is a non-vacuous separable support function with core $\mathcal{C}_b$ then there exists a *unique* collection $b_1, ..., b_n$ of non-vacuous 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) $\mathcal{C}_{b_i} \subset \mathcal{C}_b$; 4) $\mathcal{C}_{b_i} \neq \mathcal{C}_{b_j}$ if $i \neq j$.

This unique decomposition is called *canonical decomposition*. Smets [46] and Kramosil [29] solved the canonical decomposition problem by means of algebraic and measure-theoretic methods, respectively. Schubert [39] 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

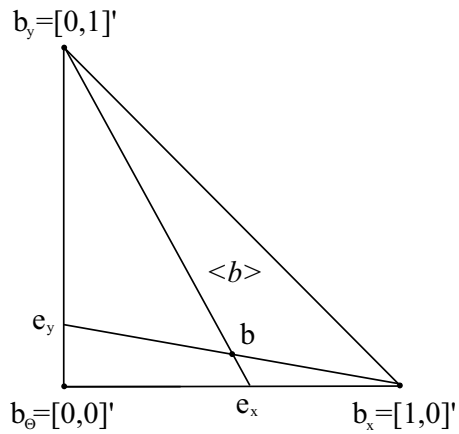

Fig. 9. Canonical decomposition of a separable support b.f. in the binary belief space.

of a separable belief function 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 Figure 9), 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.

VIII. Conclusions and perspectives

In this paper we introduced a geometric approach to the theory of evidence, in which belief functions are thought of as points of a Cartesian space whose coordinates are their belief values. Starting from the insight provided by the binary case we first proved the convexity of the belief space, and then showed that $\mathcal{B}$ has in fact the form of a *polytope* or *simplex*, in which the basic probability assignment of a b.f. b plays the role of the simplicial coordinate of b in $\mathcal{B}$. We then focused our attention on the second pillar of the ToE, Dempster's rule of combination. 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*, i.e. sets of belief functions conditioned by a given b.f. b. We concluded by discussing some open topics in the geometric approach, paying particular attention to the probabilistic approximation problem, as distances between belief functions in the belief space can be easily discussed, and approximations criteria based on Dempster's rule can be solved by exploiting the commutativity results we exposed in this paper.

A straightforward application of our characterization of the global geometry of Dempster's rule is the canonical decomposition of a belief function into simple support (pseudo) b.f., while a natural direction of research is the study of the *point-wise* behavior of $\oplus$. On the other side, a description of the geometry of possibility measures or consonant belief functions can be seen as a first step towards a unified geometric interpretations of uncertainty measures.

REFERENCES

[1] Martin Aigner, *Combinatorial theory*, Classics in Mathematics, Springer, New York, 1979.

[2] R. G. Almond, *Graphical belief modeling*, Chapman and Hall/CRC, 1995.

[3] Yaakov Bar-Shalom and Thomas E. Fortmann, *Tracking and data association*, Academic Press, Inc., 1988.

[4] P. Baroni, *Extending consonant approximations to capacities*, IPMU, 2004, pp. 1127–1134.

[5] M. Bauer, *Approximations for decision making in the Dempster-Shafer theory of evidence*, Proceedings of the Twelfth Conference on Uncertainty in Artificial Intelligence (F. Horvitz, E.; Jensen, ed.), Portland, OR, USA, 1-4 August 1996, pp. 73–80.

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

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

[8] _____, *Geometric structure of lower probabilities*, Random Sets: Theory and Applications (Goutsias, Malher, and Nguyen, eds.), Springer, 1997, pp. 361–383.

[9] B. R. Cobb and P. P. Shenoy, *A comparison of bayesian and belief function reasoning*, Information Systems Frontiers **5(4)** (2003), 345–358.

[10] _____, *A comparison of methods for transforming belief function models to probability models*, Proceedings of ECSQARU'2003, Aalborg, Denmark, July 2003, pp. 255–266.

[11] B.R. Cobb and P.P. Shenoy, *On transforming belief function models to probability models*, Tech. report, University of Kansas, School of Business, Working Paper No. 293, February 2003.

[12] F. Cuzzolin, *Geometric interplays of belief and probability*, submitted to the IEEE Transactions on Systems, Man, and Cybernetics - Part B (2005).

[13] _____, *Learning evidential models for object pose estimation*, submitted to the International Journal of Approximate Reasoning (2006).

[14] _____, *Simplicial complexes of finite fuzzy sets*, Proceedings of the 10th International Conference on Information Processing and Management of Uncertainty IPMU'04, Perugia, Italy, July 4-9, 2004, pp. 1733–1740.

[15] F. Cuzzolin and R. Frezza, *Evidential modeling for pose estimation*, Proceedings of the 4rd Internation Symposium on Imprecise Probabilities and Their Applications (ISIPTA'05), Pittsburgh, July 2005.

[16] M. Daniel, *Consistency of probabilistic transformations of belief functions*, IPMU, 2004, pp. 1135–1142.

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

[18] A. P. Dempster, *Upper and lower probability inferences based on a sample from a finite univariate population*, Biometrika **54** (1967), 515–528.

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

[20] _____, *Upper and lower probabilities inferences for families of hypothesis with monotone density ratios*, Annals of Mathematical Statistics **40** (1969), 953–969.

[21] T. Denoeux, *Inner and outer approximation of belief structures using a hierarchical clustering approach*, Int. Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **9(4)** (2001), 437–460.

[22] T. Denoeux and A. Ben Yaghlane, *Approximating the combination of belief functions using the fast moebius transform in a coarsened frame*, International Journal of Approximate Reasoning **31(1-2)** (October 2002), 77–101.

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

[24] R. Fagin and Joseph Y. Halpern, *A new approach to updating beliefs*, Uncertainty in Artificial Intelligence, 6 (L.N. Kanal P.P. Bonissone, M. Henrion and J.F. Lemmer, eds.), 1991, pp. 347–374.

[25] D. M. Gavrila and L. S. Davis, *3D model-based tracking of humans in action: A multi-view approach*, Proceedings of CVPR'96, San Francisco, CA, 18-20 June 1996, pp. 73–80.

[26] V. Ha and P. Haddawy, *Theoretical foundations for abstraction-based probabilistic planning*, Proc. of the 12th Conference on Uncertainty in Artificial Intelligence, August 1996, pp. 291–298.

[27] Va Ha and Peter Haddawy, *Geometric foundations for interval-based probabilities*, KR'98: Principles of Knowledge Representation and Reasoning (Anthony G. Cohn, Lenhart Schubert, and Stuart C. Shapiro, eds.), Morgan Kaufmann, San Francisco, California, 1998, pp. 582–593.

[28] R. Haenni and N. Lehmann, *Resource bounded and anytime approximation of belief function computations*, International Journal of Approximate Reasoning **31(1-2)** (October 2002), 103–154.

[29] Ivan Kramosil, *Measure-theoretic approach to the inversion problem for belief functions*, Fuzzy Sets and Systems **102** (1999), 363–369.

[30] H. Kyburg, *Bayesian and non-Bayesian evidential updating*, Artificial Intelligence **31:3** (1987), 271–294.

[31] Kin Chung Lo, *Agreement and stochastic independence of belief functions*, Mathematical Social Sciences **51(1)** (2006), 1–22.

[32] John D. Lowrance, T. D. Garvey, and Thomas M. Strat, *A framework for evidential-reasoning systems*, Proceedings of the National Conference on Artificial Intelligence (American Association for Artificial Intelligence, ed.), 1986, pp. 896–903.

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

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

[35] T. Moeslund and E. Granum, *A survey of computer vision-based human motion capture*, Image and Vision Computing **81** (2001), 231–268.

[36] B. Quost, T. Denoeux, and M. Masson, *One-against-all classifier combination in the framework of belief functions*, IPMU, 2006.

[37] B. Ristic and P. Smets, *Belief function theory on the continuous space with an application to model based classification*, IPMU, 2004, pp. 1119–1126.

[38] Branko Ristic and Philippe Smets, *The tbm global distance measure for the association of uncertain combat id declarations*, Information Fusion **7(3)** (2006), 276–284.

[39] Johan Schubert, *Managing decomposed belief functions*, IPMU, 2006.

[40] T. Seidenfeld, *Some static and dynamic aspects of rubust Bayesian theory*, Random Sets: Theory and Applications (Goutsias, Malher, and Nguyen, eds.), Springer, 1997, pp. 385–406.

[41] G. Shafer, *Belief functions and parametric models*, Journal of the Royal Statistical Society, Series B **44** (1982), 322–352.

[42] Glenn Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.

[43] Anna Slobodova, *Conditional belief functions and valuation-based systems*, Tech. report, Institute of Control Theory and Robotics, Slovak Academy of Sciences, Bratislava, SK, 1994.

[44] Philippe Smets, *Belief functions versus probability functions*, Uncertainty and Intelligent Systems (Saitta L. Bouchon B. and Yager R., eds.), Springer Verlag, Berlin, 1988, pp. 17–24.

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

[46] ———, *The canonical decomposition of a weighted belief*, Proceedings of the International Joint Conference on AI, IJCAI95, Montréal, Canada, 1995, pp. 1896–1901.

[47] M. Spies, *Conditional events, conditioning, and random sets*, IEEE Transactions on Systems, Man, and Cybernetics **24** (1994), 1755–1763.

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

[49] F. Voorbraak, *A computationally efficient approximation of Dempster-Shafer theory*, International Journal on Man-Machine Studies **30** (1989), 525–536.

[50] Peter Walley, *Statistical reasoning with imprecise probabilities*, Chapman and Hall, London, 1991.

[51] H. Xu and Philippe Smets, *Evidential reasoning with conditional belief functions*, Proceedings of the 10th Uncertainty in Artificial Intelligence (Lopez de Mantaras R. and Poole D., eds.), 1994, pp. 598–605.

[52] ———, *Reasoning in evidential networks with conditional belief functions*, International Journal of Approximate Reasoning **14** (1996), 155–185.

[53] A. Ben Yaghlane, T. Denoeux, and K. Mellouli, *Coarsening approximations of belief functions*, Proceedings of EC-SQARU'2001 (S. Benferhat and P. Besnard, eds.), 2001, pp. 362–373.

[54] B. Ben Yaghlane and K. Mellouli, *Belief function propagation in directed evidential networks*, IPMU, 2006.