Geometric analysis of belief space and conditional subspaces

Fabio Cuzzolin
Computer Vision Laboratory,
Università di Padova, Italy
cuzzolin@dei.unipd.it

Ruggero Frezza
Computer Vision Laboratory,
Università di Padova, Italy
frezza@dei.unipd.it

Abstract

In this paper the geometric structure of the space $\mathcal{S}_\Theta$ of the belief functions defined over a discrete set Θ (*belief space*) is analyzed. Using the Moebius inversion lemma we prove the recursive bundle structure of the belief space and show how an arbitrary belief function can be uniquely represented as a convex combination of certain elements of the fibers, giving $\mathcal{S}$ the form of a *simplex*.
The *commutativity* of orthogonal sum and convex closure operator is proved and used to depict the geometric structure of *conditional subspaces*, i.e. sets of belief functions conditioned by a given function s. Future applications of these geometric methods to classical problems like probabilistic approximation and canonical decomposition are outlined.

Keywords. Theory of evidence, belief space, fiber bundle, convex decomposition, commutativity, conditional subspace.

1 Introduction

Evidential reasoning is becoming a useful tool for solving important engineering problems. Computer vision, among the others, is an exciting field in which very difficult problems are solved by means of a wide spectrum of mathematical tools, ranging from differential geometry to statistics, to logic. When one tries and apply the theory of evidence to classical vision problems a number of important question arises.
Object tracking, for instance, consists on estimating at each time instant the current configuration of an articulated object from a sequence of images of the moving body. In a recent work of ours ([3]), image features are represented as belief functions and combined 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 parameter space $\mathcal{Q}$ of the object.
A method for deriving a pointwise estimate from the belief function emerging from the combination is then needed, for example by choosing the "best" probabilistic approximation of the belief estimate and computing the corresponding expected pose. Note that this is *not* a decision problem, and the most accurate estimate of the real vector q is desired.
Another interesting task (*data association*, [2]) consists on reconstructing the association between moving points appearing in consecutive images of a sequence. When one supposes these points belonging to an articulated body whose topological model is known, the rigid motion constraint can be used in order to achieve the desired correspondence. Since this constraint must be expressed in conditional way, the idea of combination of *conditional belief functions* (see [2], Chapter 7) in a filter-like process has to be addressed.
In this work we introduce a geometric formulation of the basis concepts of the theory of evidence, as an environment where at least some of the above problems can find a satisfactory solution. A central role is played by the notion of *belief space* $\mathcal{S}$, introduced in Section 3, while in Section 4 the Moebius inversion lemma is exploited to investigate its convexity and symmetry. With the aid of some combinatorial results, the *recursive bundle structure* of $\mathcal{S}$ is proved and an interpretation of its components (bases and fibers) in term of classes of b.f.s is provided.
Next, we characterize the relation between focal elements and the *convex closure* operator, and show in particular how every belief function can be uniquely decomposed as a convex combination of *pseudo-probabilities*, giving $\mathcal{S}$ the form of a *simplex*. In Section 6 the behavior of Dempster's rule of combination within this geometric framework is analyzed, by proving that the orthogonal sum *commutes* with the convex closure operator. This allows us to give a geometric description in term of subspaces of the collections of belief functions combinable with a given b.f. s, and the set of belief functions obtainable from s by means of combination of new evidence (*conditional subspace*).

Finally, some hints of the potential applications of this formalism are given.

1.1 Previous work

In a recent paper [7], Ha and Haddawy exploit methods of convex geometry to represent probability intervals in a computationally efficient fashion, by means of a data structure called *pcc-tree*, founded on a generalization of the convex closure called cc-operator. They show that belief functions can be represented by 2-level pcc-trees and study the evidential updating in this context.
On the other side, a number of papers have been published on approximation of belief functions (see [1] for a review), mainly in order to find efficient implementations of the rule of combination aiming to reduce the number of focal elements. Tessem ([15]) incorporates only the highest-valued focal elements in his m_{klx} approximation; a similar approach inspires the *summarization* technique formulated by Lowrance *et al.* ([10]).
About conditional belief functions, Fagin and Halpern ([6]) formulate a definition based on inner measures, as lower envelope of a family of conditional probability functions, and provide a closed-form expression for it. M. Spies ([14]) establishes a link between conditional events and discrete random sets, defining conditional events as sets of equivalent events under the conditioning relation. By applying to them a multivalued mapping ([4]), he defines conditional belief functions and introduces an updating rule (that is equivalent to the law of total probability is all beliefs are probabilities).

2 Belief functions

Following Shafer [11] we call a finite set of possibilities *frame*[1] *of discernment* (FOD).

Definition 1. *A* basic probability assignment *(b.p.a.) over a FOD Θ is a function $m : 2^\Theta \rightarrow [0,1]$ such that*

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

The elements of 2^Θ associated to non-zero values of m are called *focal elements* and their union *core*. Now suppose a b.p.a. is introduced over an arbitrary FOD.

Definition 2. *The* belief function *Bel associated to a basic probability assignment* m *is defined as:*

$$Bel(A) = \sum_{B \subset A} m(B).$$

An alternative definition of belief functions can be given independently from the presence of a basic probability assignment (see [11]).
Belief functions representing distinct bodies of evidence are combined by means of the *Dempster's rule of combination.*

Definition 3. *the* orthogonal sum *$Bel_1 \oplus Bel_2$ of two belief functions is a function whose focal elements are all the possible intersections between the combining focal elements and whose b.p.a. is given by*

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

The normalization constant in the above expression measures the *level of conflict* between belief functions for it represents the amount of probability they attribute to contradictory (i.e. disjoint) subsets.
Dempster's rule is easily extended to the combination of several belief functions.

In the theory of evidence a probability function is simply a peculiar belief function satisfying the additivity rule for disjoint sets.

Definition 4. *A* Bayesian *belief function assigns basic probabilities only to singletons $\theta \in \Theta$ of the underlying frame: $m(A) = 0\ \ \forall\ A$ s.t. $|A| > 1$.*

It can be proved that

Proposition 1. *A function Bel is Bayesian $\Leftrightarrow \exists\ p : \Theta \rightarrow [0,1]$ such that*

$$\sum_{\theta \in \Theta} p(\theta) = 1,\ Bel(A) = \sum_{\theta \in A} p(\theta)\quad \forall A \subset \Theta.$$

3 Belief space

Consider a frame of discernment Θ and introduce in the Euclidean space $\mathbb{R}^{|2^\Theta|}$ an orthonormal reference frame $\{x_i\}_{i=1,\ldots,|2^\Theta|}$ in which, given an arbitrary ordering in 2^Θ, each coordinate function x_i measures the value of belief associated to a the i-th subset of Θ.

Definition 5. *The* belief space *associated to Θ is the set of points $\mathcal{S}$ of $\mathbb{R}^{|2^\Theta|}$ corresponding to a belief function.*

3.1 Limit simplex

The properties of Bayesian belief functions can be useful to have a first idea of the shape of the belief space.

Lemma 1. *If p is a Bayesian belief function over a frame Θ and B an arbitrary subset of Θ*

$$\sum_{A \subset B} p(A) = 2^{|B|-1} \cdot P(B).$$

Proof. The sum can be rewritten as $\sum_{\theta \in B} k_\theta p(\theta)$ where k_θ is the number of subsets of B containing θ. But $k_\theta = 2^{|B|-1}$ for each singleton, so that

$$\sum_{A \subset B} p(A) = 2^{|B|-1} \cdot \sum_{\theta \in B} p(\theta) = 2^{|B|-1} \cdot P(B).$$

$\square$

As a consequence all the Bayesian functions are constrained to belong to a well-determined region of the belief space.

Corollary 1. *The set $\mathcal{P}$ of all the Bayesian belief functions with domain Θ is included into the $|\Theta| - 1$-dimensional simplex*

$$\mathcal{L} = \{s \ s.t. \sum_{A \subset \Theta} s(A) = 2^{|\Theta|-1}\}$$

of the $|\Theta|$-dimensional belief space $\mathcal{S}$, called limit simplex.

Theorem 1. *The set of all the belief functions over a frame Θ is a subset of the region delimited by the limit simplex $\mathcal{L}$:*

$$\sum_{A \subset \Theta} s(A) \leq 2^{|\Theta|-1}$$

where the equality holds iff s is Bayesian.

Proof. The sum can be written as

$$\sum_{i=1}^{f} a_i \cdot m(A_i)$$

where f is the number of focal elements of s and a_i is the number of subsets of Θ including the $i_t h$ focal element A_i, namely $a_i = |\{B \subset \Theta \ s.t. \ B \supset A_i\}|$. Obviously $a_i = 2^{|\Theta \setminus A|} \leq 2^{|\Theta|-1}$ and the equality holds iff $|A| = 1$. Then

$$\sum_{A \subset \Theta} s(A) = \sum_{i=1}^{f} m(A_i) \cdot 2^{|\Theta|-1}$$

$$= 2^{|\Theta|-1} \cdot \sum_{i=1}^{f} m(A_i) = 2^{|\Theta|-1} \cdot 1 = 2^{|\Theta|-1}$$

iff $|A_i| = 1$ for every focal element of s, i.e. s is Bayesian. $\square$

It is important to point out that $\mathcal{P}$ in general does not sell out the limit simplex $\mathcal{L}$. At the same time the belief space generally does not coincide with the entire region bounded by $\mathcal{L}$, as is shown by the picture of the belief space $\mathcal{S}_2$ associated to a binary frame.

Example. Suppose $\Theta = \{\theta_1, \theta_2\}$: we have $\mathcal{P} = \mathcal{L} \cap \{s : s(\Theta) = 1\}$ and the belief space has the form illustrated in Figure 1.

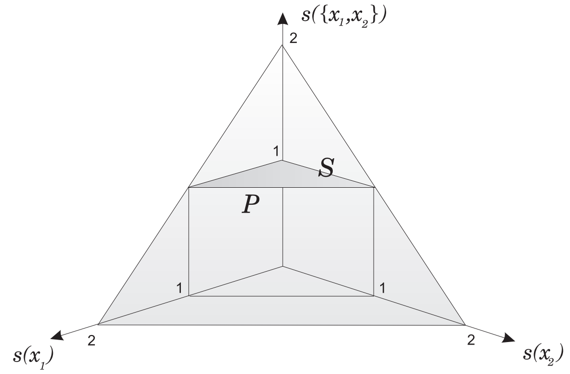

Figure 1: The belief space $\mathcal{S}$ and its probabilistic border $\mathcal{P}$ for a binary frame.

3.2 Upper probabilities

Another hint on the structure of $\mathcal{S}$ comes from the particular relation of Bayesian belief functions with the classical L_1 distance in the Euclidean space.
Let $\mathcal{C}_s$ denote the core of s, and define the following order relation:

$$s \geq s' \equiv s(A) \geq s'(A) \quad \forall A \subset \Theta.$$

Lemma 2. *If $s \geq s'$ then $\mathcal{C}_s \subset \mathcal{C}_{s'}$.*

Proof. Obviously, since $s(A) \geq s'(A)$ for every $A \subset \Theta$, that is true for $\mathcal{C}_{s'}$ too, i.e. $s(\mathcal{C}_{s'}) = 1$ but then $\mathcal{C}_s \subset \mathcal{C}_{s'}$. $\square$

Theorem 2. *If $s : 2^\Theta \to [0, 1]$ is an arbitrary belief function over a frame Θ then*

$$\|s - p\|_{L_1} = \sum_{A \subset \Theta} |s(A) - p(A)| = cost = f(s)$$

for every Bayesian function $p : 2^\Theta \to [0, 1]$ which is greater then s, i.e.

$$p(A) = \sum_{\theta \in A} p(\{\theta\}) \geq s(A) \quad \forall A \subset \Theta.$$

Proof. Lemma 2 guarantees that $\mathcal{C}_p \subset \mathcal{C}_s$, so that $p(A) - s(A) = 1 - 1 = 0$ for $A \supset \mathcal{C}_s$. On the other hand, if $A \cap \mathcal{C}_s = \emptyset$ then $p(A) - s(A) = 0 - 0 = 0$. What is left are sets corresponding to unions of non empty proper subsets of $\mathcal{C}_s$ and arbitrary subsets of $\Theta \setminus \mathcal{C}_s$. Given $A \subseteq \mathcal{C}_s$ there are $2^{|\Theta \setminus \mathcal{C}_s|}$ subsets of the above type containing it, so that

$$\sum_{A \subset \Theta} |s(A) - p(A)| = 2^{|\Theta \setminus \mathcal{C}_s|} \cdot [\sum_{A \subseteq \mathcal{C}_s} p(A) - \sum_{A \subseteq \mathcal{C}_s} s(A)]$$

but then for Lemma 1

$$= 2^{|\Theta \setminus \mathcal{C}_s|} \cdot [2^{|\mathcal{C}_s|-1} - 1 - \sum_{A \subseteq \mathcal{C}_s} s(A)]$$

that is the size $f(s)$ of the upper probability simplex and is independent from p. $\square$

A probability distribution satisfying the hypothesis of Theorem 2 is said to be *consistent* with s ([9]). Ha *et al.* ([7]) proved that the set $P(s)$ of probability functions consistent with a given b.f. s can be expressed (in the *probability* simplex, not the belief space) as the sum of the probability simplexes associated to its focal elements A_i, $i = 1, ..., k$ of s, weighted by their masses:

$$P(s) = \sum_{i=1}^{k} m(A_i) conv(A_i)$$

where $conv(A_i)$ is the convex closure of the probabilities $\{P_\theta | \theta \in A_i\}$ assigning 1 to an element of A_i.

4 Bundle structure

These preliminary results suggest the belief space should have the form of a simplex. A more detailed description needs resorting to the axioms of basic probability assignments (see Definition 1).

4.1 Moebius inversion lemma

Given a belief function s, the corresponding basic probability assignment can be found by applying the so-called *Moebius inversion lemma*

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

that comes out from the structure of poset of power sets. We can exploit it to determine whether a point $s \in \mathbb{R}^{|2^\Theta|}$ corresponds to a belief function, by simply computing the b.p.a. and checking the axioms m must obey.
The *normalization* constraint $\sum_{A \subset \Theta} m(A) = 1$ trivially translates into $\mathcal{S} \subset \{s : s(\Theta) = 1\}$. The *positivity* condition is more interesting, for it originates an inequality resounding the third axiom of belief functions ([11], page 5). $\forall A \subset \Theta$:

$$\begin{aligned} & s(A) - \textstyle\sum_{B \subset A, |B|=|A|-1} s(B) + ... \\ & ... + (-1)^{|A-B|} \textstyle\sum_{|B|=k} s(B) + ... \\ & ... + (-1)^{|A|-1} \textstyle\sum_{\theta \in \Theta} s(\{\theta\}) \geq 0. \end{aligned} \tag{2}$$

4.1.1 Example: ternary frame

Let us see the form of the belief space in a simple but significant case: the ternary frame $\Theta = \{\theta_1, \theta_2, \theta_3\}$. If we denote the coordinates with

$$x = s(\{\theta_1\}),\ y = s(\{\theta_2\}),\ z = s(\{\theta_3\}),\ u = s(\{\theta_1, \theta_2\}),$$
$$v = s(\{\theta_1, \theta_3\}),\ w = s(\{\theta_2, \theta_3\})$$

the positivity constraint (2) can be rewritten as

$$\mathcal{S} : \begin{cases} x \geq 0, \ \ u \geq (x + y) \\ y \geq 0, \ \ v \geq (x + z) \\ z \geq 0, \ \ w \geq (y + z) \\ 1 - (u + v + w) + (x + y + z) \geq 0 \end{cases} \tag{3}$$

By combining the last equation in (3) with the others we obtain

$$0 \leq x + y + z \leq 1, \ \ 0 \leq u + v + w \leq 2;$$

called $k = x + y + z$, it necessary follows that

$$2k \leq u + v + w \leq 1 + k$$

$$u \geq (x + y), \ v \geq (x + z), \ w \geq (y + z).$$

In other words, $\mathcal{S}$ reveals the structure of *fiber bundle*, in which the variables x, y, z can move freely in the unitary simplex, while the others are constrained to stay in a tetrahedron R_P that depends on the sum $x + y + z = k$ (see Figure 3).
System (3) shows a natural symmetry that reflects the

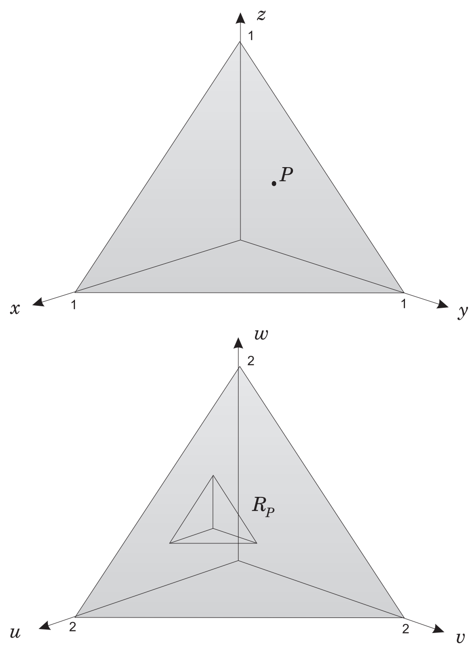

Figure 2: Decomposition of the belief space in the ternary case.

intuitive partition of the variables in two sets, each associated to subsets of Θ with a same size, respectively $\{x, y, z\} \sim |A| = 1$ and $\{u, v, w\} \sim |A| = 2$.
It is easy to see that the group of symmetry of $\mathcal{S}$ is the permutation group $S(3)$, acting onto $\{x, y, z\} \times \{u, v, w\}$ by means of the correspondence

$$x \leftrightarrow w, \quad y \leftrightarrow v, \quad z \leftrightarrow u.$$

4.2 Convexity

All the constraints in Equation (2) defining $\mathcal{S}$ are of the form

$$\sum_{i \in G_1} x_i \geq \sum_{j \in G_2} x_j$$

where G_1 and G_2 are two disjoint sets of coordinates, as the above example confirms. As a straightforward consequence,

Theorem 3. *$\mathcal{S}$ is convex.*

Proof. Let us take two points $P_0, P_1 \in \mathcal{S}$ and prove that all the points of the segment $P_0 + \alpha(P_1 - P_0)$, $0 \leq \alpha \leq 1$, belong to $\mathcal{S}$. Since $P_0,\ P_1 \in \mathcal{S}$

$$\sum_{i \in G_1} x_i^0 \geq \sum_{j \in G_2} x_j^0, \quad \sum_{i \in G_1} x_i^1 \geq \sum_{j \in G_2} x_j^1$$

where x_i^0, x_i^1 are the i-th coordinates of P_0, P_1 respectively, so that

$$\begin{aligned}
&\textstyle\sum_{i \in G_1} x_i^\alpha = \sum_{i \in G_1} [x_i^0 + \alpha(x_i^1 - x_i^0)] \\
&= \textstyle\sum_{i \in G_1} x_i^0 + \alpha \sum_{i \in G_1} (x_i^1 - x_i^0) \\
&= \textstyle(1 - \alpha) \sum_{i \in G_1} x_i^0 + \alpha \sum_{i \in G_1} x_i^1 \geq \\
&\geq \textstyle(1 - \alpha) \sum_{j \in G_2} x_j^0 + \alpha \sum_{j \in G_2} x_j^1 \\
&= \textstyle\sum_{j \in G_2} [x_j^0 + \alpha(x_j^1 - x_j^0)] = \sum_{j \in G_2} x_j^\alpha.
\end{aligned}$$

□

It is well-known that belief functions are a special type of *coherent lower probabilities*, that in turn can be seen as a particular class of *lower previsions* (consult [17], Section 5.13). Walley has proved that coherent lower probabilities are closed under convex combination; this implies that convex combinations of belief functions (completely monotone lower probabilities) are still coherent. Theorem 3 is a stronger result, stating that they are also completely monotone.

4.3 Symmetry

The above remark about the symmetry of the belief space in Example 4.1.1 can be extended to the general case of a finite n-dimensional frame $\Theta = \{\theta_1, ..., \theta_n\}$. Let us establish for sake of simplicity the following notation: $x_i x_j ... x_k \doteq s(\{\theta_i, \theta_j, ..., \theta_k\})$.

Proposition 2. *The general symmetry of the belief space is described by the following logic expression*

$$\bigvee_{1 \leq i,j \leq n} \ \bigwedge_{k=1}^{n-1} \ \bigwedge_{\substack{\{i_1, ..., i_{k-1}\} \subset \\ \subset \{1, ..., n\} \setminus \{i, j\}}} \begin{array}{c} x_i x_{i_1} \cdots x_{i_{k-1}} \\ \updownarrow \\ x_j x_{i_1} \cdots x_{i_{k-1}} \end{array}$$

where $\bigvee (\bigwedge)$ denotes the logical or (and), while $\leftrightarrow$ indicates the permutation of pairs of coordinates.

Proof. Let us rewrite the Moebius constraints, using the above notation:

$$x_{i_1} \cdots x_{i_k} \geq \sum_{l=1}^{k-1} (-1)^{k-l+1} \sum_{\{j_1, ..., j_l\} \subset \{i_1, ..., i_k\}} x_{j_1} \cdots x_{j_l}$$

Looking at the right side of the equation, it is clear that only a permutation between coordinates associated to subsets of the *same size* may leave the inequality inalterate.
Given the triangular form of the system of inequalities (the first group concerning variables of size 1, the second one variables of size 1 and 2, and so on), permutations among size k variables are always induced by permutations of variables of smaller size. Hence the symmetries of $\mathcal{S}$ are determined by permutations of singletons. But such an exchange $x_i \leftrightarrow x_j$ determines a sequence of permutations among the coordinates related to subsets containing θ_i and θ_j.
The resulting symmetry V_k induced by $x_i \leftrightarrow x_j$ for the k-th group of constraints is then

$$(x_i \leftrightarrow x_j) \wedge \cdots \wedge (x_i x_{i_1} \cdots x_{i_{k-1}} \leftrightarrow x_j x_{i_1} \cdots x_{i_{k-1}})$$

$$\forall \{i_1, ..., i_{k-1}\} \subset \{1, ..., n\} \setminus \{i, j\}.$$

Since V_k is obviously implied by V_{k+1}, and V_n is always trivial (as a simple check confirms), the overall symmetry induced by a permutation of singletons is determined by V_{n-1}, and by considering all the possible permutations $x_i \leftrightarrow x_j$ we have the thesis. □

In other words, the symmetry of $\mathcal{S}$ is determined by the action of the permutation group $S(n)$ over the collection of the size 1 variables *and* the action of $S(n)$ naturally induced on the other variables

$$s \in S(n) : \quad \begin{array}{ccc} P_k(\Theta) & \to & P_k(\Theta) \\ x_{i_1} \cdots x_{i_k} & \mapsto & s x_{i_1} \cdots s x_{i_k} \end{array}$$

where $P_k(\Theta)$ is the class of the size k subsets of Θ. It is not difficult to recognize the symmetry properties of a *simplex*, i.e. a collection $[\sigma_1, ..., \sigma_n]$ of points in the Euclidean space together with the sub-collections (faces) $[\sigma_{i_1}, ..., \sigma_{i_k}]$ of all the orders $k \leq n$.

4.4 Recursive bundle structure

The decomposition property shown in Example 4.1.1 is a hint of a general feature of the belief space: it can be *recursively decomposed into fibers* parameterized by the coordinates related to subsets $A \subset \Theta$ with a same size. We first need some simple combinatorial results.

Lemma 3.

$$\sum_{m=l}^{k-1}(-1)^{m-l}\binom{n-l}{m-l} = (-1)^{k-l+1}\binom{n-(l+1)}{k-(l+1)}.$$

Proof.

$$\sum_{m=l}^{k-1}(-1)^{m-l}\binom{n-l}{m-l} = \binom{n-l}{0} - \binom{n-l}{1} + ...$$

$$... + (-1)^{k-(l+1)}\binom{n-l}{k-(l+1)} =$$

$$= \frac{(k-(l+1))! + ... + (-1)^{k-(l+1)}(n-l)\cdot ... \cdot(n-k+2)}{(k-(l+1))!}$$

The first two terms of the numerator have a common factor and can be rewritten as $-(k-(l+1))!\cdot(n-l-1)$; this in turn has in common $(n-l-1)(k-(l+1))!/2!$ with the third addendum giving $(n-l-1)(n-l-2)(k-(l+1))!/2!$ and so on. By iteration we get the thesis. □

Theorem 4.

$$\sum_{|A|=k} s(A) \leq 1 + \sum_{m=1}^{k-1}(-1)^{k-m+1}\binom{n-(m+1)}{k-m}\cdot\sum_{|B|=m} s(B).$$

Proof. Let us suppose we have assigned an amount of mass to the subsets of size smaller than k, $\{m(B), |B| < k\}$ with $\sum_{|B|<k} m(B) < 1$. The maximum value of $\sum_A s(A)$ corresponds to assigning the remaining mass $1-\sum_{|A|<k} m(A)$ to the subsets of size k. We obtain

$$1 - \sum_{|A|<k} m(A) = 1 - \sum_{|A|<k}\sum_{B\subset A}(-1)^{|A-B|}s(B)$$

$$= 1 - \sum_{|A|=m=1}^{k-1}\sum_{|B|=l=1}^{m}(-1)^{m-l}s(B)$$

$$= 1 - \sum_{|A|=m=1}^{k-1}\sum_{|B|=l=1}^{m}(-1)^{m-l}\binom{n-l}{m-l}\cdot\sum_{|B|=l}s(B)$$

for $\binom{n-l}{m-l}$ is the number of subsets of size m containing a fixed set B, $|B| = l$ in a frame with n elements. This can be rewritten as

$$1 - \sum_{|B|=l=1}^{k-1}(\sum_{|B|=l}s(B))\cdot\sum_{m=l}^{k-1}(-1)^{m-l}\binom{n-l}{m-l}$$

that using Lemma 3 becomes

$$1 - \sum_{|B|=l=1}^{k-1}(\sum_{|B|=l}s(B))\cdot(-1)^{k-l+1}\binom{n-(l+1)}{k-(l+1)}. \quad (4)$$

Now, if we write $\sum_{|A|=k} s(A) =$

$$= \sum_{|A|=k}\sum_{B\subset A} m(A) = \sum_{l=1}^{k}\sum_{|B|=l} m(B)\cdot\binom{n-l}{k-l} =$$

(since $\binom{n-l}{k-l}$ is the number of size k subsets including size i subsets)

$$= \sum_{|B|=k} m(B) + \sum_{l=1}^{k-1}\binom{n-l}{k-l}\cdot(\sum_{|B|=l} m(B))$$

$$= \sum_{|B|=k} m(B) +$$

$$+\sum_{l=1}^{k-1}\binom{n-l}{k-l}\cdot[\sum_{m=1}^{l}(-1)^{l-m}\binom{n-m}{l-m}\sum_{|B|=m}s(B)]$$

$$= \sum_{|B|=k} m(B) + \sum_{l=1}^{k-1}(-1)^{k-l-1}\binom{n-l}{k-l}\sum_{|B|=l}s(B);$$

by substituting to the first addendum Equation (4) we get

$$= 1 + \sum_{l=1}^{k-1}(-1)^{k-l-1}\cdot(\sum_{|B|=l}s(B))\cdot[\binom{n-l}{k-l} - \binom{n-(l+1)}{k-(l+1)}]$$

and by developing the pair of binomials we have the thesis. □

Now, let us recall the notation used in Proposition 2.

Theorem 5. *The belief space $\mathcal{S}$ has a recursive bundle structure, i.e. it can be decomposed into the following sequence of domains*

$$\mathcal{S} \xrightarrow{p_1} \mathcal{D}^{(1)},$$

$$\mathcal{F}^{(1)} \doteq p_1^{-1}(x_1, ..., x_n) = \mathcal{S}_{x_i}$$

$$\mathcal{F}^{(1)} \xrightarrow{p_2} \mathcal{D}^{(2)},$$

$$\mathcal{F}^{(2)} \doteq p_2^{-1}(x_1x_2, ..., x_{n-1}x_n)) = \mathcal{S}_{x_i, x_ix_j}$$

$$\cdots$$

$$\mathcal{F}^{(n-2)} \xrightarrow{p_{n-1}} \mathcal{D}^{(n-1)},$$

$$\mathcal{F}^{(n-1)} \doteq p_{n-1}^{-1}(x_1\cdots x_{n-1}, ..., x_2\cdots x_n)) =$$
$$= \mathcal{S}_{x_i, ..., x_{i_1}\cdots x_{i_{n-1}}}$$

where $\mathcal{S}_{x_i,...,x_{i_1}\cdots x_{i_k}}$ is the section of the belief space with $x_{i_1}\cdots x_{i_j}$ =cost $\forall\ \{i_1,...,i_j\}\subset\{1,...,n\}$, $\forall j\leq k$ (i.e. the set of all the possible belief functions whose basic probability assignment $m(A)$ is fixed for $|A|\leq k$).
The i-th basis *of $\mathcal{S}$, $\mathcal{D}^{(i)}$, is defined by the following equations*

$$x_{k_1}\cdots x_{k_i} \geq \sum_{m<i}(-1)^{i-m+1} \sum_{\substack{\{l_1,...,l_m\}\subset\\ \subset\{k_1,...,k_i\}}} x_{l_1}\cdots x_{l_m}$$

$$x_{k_1}\cdots x_{k_j} = \sum_{m<j}(-1)^{j-m+1} \sum_{\substack{\{l_1,...,l_m\}\subset\\ \subset\{k_1,...,k_j\}}} x_{l_1}\cdots x_{l_m}$$

$$\sum_{\substack{\{k_1,...,k_i\}\subset\\ \subset\{1,...,n\}}} x_{k_1}\cdots x_{k_i} \leq 1+\sum_{m=1}^{i-1}(-1)^{i-m+1}\cdot \binom{n-(m+1)}{i-m}\cdot \sum_{\substack{\{k_1,...,k_m\}\subset\\ \subset\{1,...,n\}}} x_{k_1}\cdots x_{k_m}$$

with $i<j\leq n$. $\mathcal{F}^{(i)}$ is called the the i-th fiber *of the belief space, while p_i is the projection map of the i-th bundle level.*

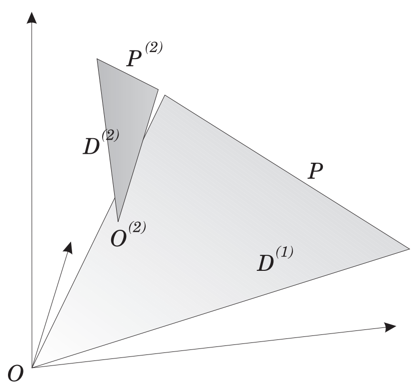

Figure 3: Pictorial representation of the bundle structure of the belief space.

Remark. Note that $\mathcal{D}^{(i)}$ is $\binom{n}{i}$-dimensional and is parameterized by the variables $x_{k_1}\cdots x_{k_i}$ associated to the size i subsets of Θ, since the values of the high-order variables are bond to them by the second group of equations.

Proof. (sketch) It suffices to observe that the first group of constraints defining $\mathcal{D}^{(i)}$ summarizes the Moebius inequalities for subsets of size i, while the second one means that all the subsets with size greater than i satisfy the Moebius formulae as equalities. The second equation comes directly from Theorem 4. $\square$

Remark. $D^{(i)} = Cl(O^{(i)}, P^{(i)})$, where $P^{(i)}$ is the collection of belief functions assigning all the remaining basic probability to subsets of size i, while $O^{(i)}$ assigns all the mass to Θ.

Figure 3 summarizes our knowledge of the bundle structure of the belief space. $\mathcal{S}$ can be decomposed into a base $\mathcal{D}^{(1)}$ (the simplex $u=(x+y)$, $v=(x+z)$, $w=(y+z)$ in Example 4.1.1) whose points are glued to a fiber (R_P in the ternary case) whose dimension reduces to zero at the upper border $\mathcal{P}^{(1)}$ of $\mathcal{D}^{(1)}$. This decomposition recursively applies to the fibers, for $i=1,...,n-1$.
It is interesting to point out that the elements of this decomposition have an intuitive meaning. For instance, $\mathcal{P}^{(1)}=\mathcal{P}$ is the set of the Bayesian belief functions, while $\mathcal{D}^{(1)}$ coincides to the collection of the *discounted* probabilities (see [11]).

5 Simplicial form of the belief space

The bundle structure of the belief space introduced above coexists with a simpler representation, resounding the polytope-like description of the set of probability distributions on a given domain ([7]).

Definition 6. *The set* $\Pi^{(i)}$ *of the* pseudo-probabilities *of order i is the collection of belief functions assigning all the basic probability to subsets of size i.*

Theorem 6. *Every belief function $s\in\mathcal{S}$ can be* uniquely *expressed as a convex combination of pseudo-probabilities of all the orders,*

$$s=\sum_{i=1}^{n}\alpha_i\pi^{(i)},\qquad \sum_{i=0}^{n}\alpha_i=1,\qquad \pi^{(i)}\in\Pi^{(i)}.$$

Proof.

$$s=(\sum_{B\subset A}m(B),\ A\subset\Theta)=\sum_{i=1}^{n}(\sum_{B\subset A,|B|=i}m(B),\ A\subset\Theta)$$

but then

$$\delta^{(i)}\doteq(\sum_{B\subset A,|B|=i}m(B),\ A\subset\Theta)$$

is an unnormalized belief function assigning all the mass to subsets of size i.
By defining $\pi^{(i)}\doteq\frac{\delta^{(i)}}{\|\delta^{(i)}\|}$ we can write

$$s=\sum_{i=1}^{n}\|\delta^{(i)}\|\pi^{(i)}=\sum_{i=1}^{n}\alpha_i\pi^{(i)}$$

with $\sum_{i=1}^{n}\alpha_i=\sum_{i=1}^{n}\|\delta^{(i)}\|=1$ for the normalization constraint. Obviously $\pi^{(i)}\in\Pi^{(i)}$. $\square$

Not surprisingly, $\Pi^{(i)}=P_0^{(i)}$.
This convex decomposition property can be easily generalized in the following way.

Theorem 7. *The set of all the belief functions with focal elements in a given collection X is closed and convex in $\mathcal{S}$, namely*

$$\{s : A \in \mathcal{E}_s \Rightarrow A \in X\} = Cl(\{P_A : A \in X\})$$

where P_A is the pseudo-probability assigning all the mass to A, $m_{P_A}(A) = 1$.

Proof. By definition $\{s : A \in \mathcal{E}_s \Rightarrow A \in X\} =$

$$= \{s : s = (\sum_{B\subset A, B\in\mathcal{E}_s} m(B)), \mathcal{E}_s \subset X\}$$

but

$$s = \sum_{B\in\mathcal{E}_s} m(B) \cdot (x_A = 1, A \supset B; x_A = 0, A \not\supset B)$$

and $(x_A = 1, A \supset B; x_A = 0, A \not\supset B) \doteq P_B$, so that

$$s = \sum_{B\in\mathcal{E}_s} m(B)P_B = \sum_{B\in X} m(B)P_B$$

by extending m to the elements $B \in X \setminus \mathcal{E}_x$ as $m(B) = 0$. Since m is a basic probability assignment, $\sum_{B\in X} m(B) = 1$ and the thesis follows. □

Corollary 2. $\Pi^{(i)} = Cl(\{P_B,\ |B| = i\})$.

Corollary 3. *The belief space $\mathcal{S}$ coincides to the convex closure of all the pseudo-probabilities of every order,*

$$\mathcal{S} = Cl(\Pi^1, ..., \Pi^{n-1}, \underline{0}) = Cl(P_1^1, ..., P_n^1, ..., P_x^{n-1}, \underline{0}).$$

The above result (that can be obtained directly from Theorem 7) confirms our conjecture about the nature of simplex of the belief space induced by the symmetry analysis of Paragraph 4.3.

6 Commutativity

Once established the geometrical properties of the belief functions, it is natural to wonder what is the behaviour of the rule of combination in the framework of the belief space.

Theorem 8. *Cl and $\oplus$ commute, i.e. if s is combinable with s_i, $\forall i = 1, ..., n$ then*

$$s \oplus Cl(\{s_1\}, ..., \{s_n\}) = Cl(\{s \oplus s_1\}, ..., \{s \oplus s_n\});$$

in other words

$$s \oplus \sum_i \alpha_i s_i = \sum_i \alpha_i (s \oplus s_i), \qquad \sum_i \alpha_i = 1.$$

Remark. Being $\mathcal{S}$ convex, if $s_i \in \mathcal{S}\ \forall i$ then $\sum_i \alpha_i s_i \in \mathcal{S}$ when $\sum_i \alpha_i = 1$.

Proof. Let us first compute the basic probability assignment associated to $\sum_i \alpha_i s_i$, by means of the Moebius inversion formula (1). If by hypothesis $s(B) = \sum_i \alpha_i s_i(B)$ then

$$m_{\sum_i \alpha_i s_i}(A) = \textstyle\sum_{B\subset A} (-1)^{|A-B|} \sum_i \alpha_i s_i(B) =$$

$$\textstyle\sum_i \alpha_i \cdot \sum_{B\subset A} (-1)^{|A-B|} s_i(B) = \sum_i \alpha_i m_i(A).$$

Now, being $\sum_i \alpha_i s_i \in \mathcal{S}$ for the above remark, we must check if it combinable with with s, obtaining $s \oplus \sum_i \alpha_i s_i$. Called $\mathcal{E}_s$ the collection of focal elements of a belief function s, we have

$$\mathcal{E}_{\sum_i \alpha_i s_i} = \bigcup_{i:\alpha_i\neq 0} \mathcal{E}_{s_i}; \qquad (5)$$

if $\alpha_i \neq 0\ \forall\ i$ this reduces to $\mathcal{E}_{\sum_i \alpha_i s_i} = \bigcup \mathcal{E}_{s_i}$. This way if s is combinable with some s_i (even only one of them) then it is combinable with $\sum_i \alpha_i s_i$.
Let us call $A_1, ..., A_n$ the focal elements of $\sum_i \alpha_i s_i$ and $B_1, ..., B_m$ those of s. The f.e. of $s \oplus \sum_i \alpha_i s_i$ are

$$\bigcup_i \mathcal{E}_{s\oplus s_i}$$

for all the intersections are considered, but Property (5) gives exactly the same result for the f.e. of $\sum_i \alpha_i (s \oplus s_i)$. Hence, we have to check the corresponding basic probability assignments: for the latter we have, denoting with $\{E_k\}$ the focal elements of s_i,

$$m_{\sum_i \alpha_i s\oplus s_i}(A) = \textstyle\sum_{B\subset A} (-1)^{|A-B|} \sum_i \alpha_i (s \oplus s_i)(B)$$

$$= \textstyle\sum_i \alpha_i \sum_{B\subset A} (-1)^{|A-B|} (s \oplus s_i)(B) =$$

$$\textstyle\sum_i \alpha_i m_{s\oplus s_i}(A) = \sum_i \alpha_i \frac{\displaystyle\sum_{E_k\cap B_j = A} m_{s_i}(E_k) m_s(B_j)}{\displaystyle 1 - \sum_{E_k\cap B_j=\emptyset} m_{s_i}(E_k) m_s(B_j)}$$

while, for $s \oplus \sum_i \alpha_i s_i$,

$$m_{s\oplus\sum_i \alpha_i s_i}(A) = \frac{\displaystyle\sum_{A_k\cap B_j=A} m_{\sum_i \alpha_i s_i}(A_k) m_s(B_j)}{\displaystyle 1 - \sum_{A_k\cap B_j=\emptyset} m_{\sum_i \alpha_i s_i}(A_k) m_s(B_j)}$$

$$= \frac{\displaystyle\sum_{A_k\cap B_j=A} (\sum_i \alpha_i m_{s_i}(A_k)) \cdot m_s(B_j)}{\displaystyle 1 - \sum_{A_k\cap B_j=\emptyset} (\sum_i \alpha_i m_{s_i}(A_k)) \cdot m_s(B_j)}$$

$$= \frac{\displaystyle\sum_i \alpha_i \cdot \sum_{A_k\cap B_j=A} m_{s_i}(A_k) m_s(B_j)}{\displaystyle\sum_i \alpha_i - \sum_{A_k\cap B_j=\emptyset} (\sum_i \alpha_i m_{s_i}(A_k)) \cdot m_s(B_j)}$$

$$= \frac{\sum_i \alpha_i \cdot \sum_{A_k \cap B_j = A} m_{s_i}(A_k) m_s(B_j)}{\sum_i \alpha_i \cdot (1 - \sum_{A_k \cap B_j = \emptyset} m_{s_i}(A_k) m_s(B_j))}$$

$$= \sum_i \alpha_i \cdot \frac{\sum_{A_k \cap B_j = A} m_{s_i}(A_k) m_s(B_j)}{1 - \sum_{A_k \cap B_j = \emptyset} m_{s_i}(A_k) m_s(B_j)}.$$

Since for $A_k \notin \mathcal{E}_{s_i}$ the addenda vanish, we remain for each i with the focal elements of s_i:

$$\sum_i \alpha_i \cdot \frac{\sum_{E_k \cap B_j = A} m_{s_i}(E_k) m_s(B_j)}{1 - \sum_{E_k \cap B_j = \emptyset} m_{s_i}(E_k) m_s(B_j)}, \quad E_k \in \mathcal{E}_{s_i}.$$

□

The fact that the orthogonal sum and convex closure operators commute is a powerful tool. It provides a simple language that allows us to give geometric interpretations of the notions of combinability and conditioning.

7 Conditional subspaces

Definition 7. *The* conditional subspace $\langle s \rangle$ *is the set of all the belief functions conditioned by a given function s, namely*

$$\langle s \rangle \doteq \{s \oplus t,\ t \in \mathcal{S}\ s.t.\ \exists\ s \oplus t\}. \tag{6}$$

Since not every belief function is combinable with an arbitrary s, we need to understand the geometric structure of combinable functions.

Definition 8. *The* non-combinable subspace $NC(s)$ *associated to a belief function s is the collection of all the b.f.s not combinable with s,*

$$NC(s) \doteq \{s' : \nexists s' \oplus s\}.$$

Proposition 3. $NC(s) = Cl(\{P_A : A \cap \mathcal{C}_s = \emptyset\})$.

Proof. It suffices to point out that $NC(s) = \{s' : \mathcal{C}_{s'} \subset \overline{\mathcal{C}}_s\} = \{s' : A \subset \overline{\mathcal{C}}_s \ \ \forall A \in \overline{\mathcal{C}}_{s'}\}$. Hence we can apply Theorem 7 and the thesis follows. □

The dimension of $NC(s)$ is obviously $2^{|\Theta \setminus \mathcal{C}_s|} - 2$.
Using the definition of non-combinable subspace we can write $\langle s \rangle = s \oplus (\mathcal{S} \setminus NC(s)) = s \oplus \{s' : \mathcal{C}_{s'} \cap \mathcal{C}_s \neq \emptyset\}$. Unfortunately, the last expression does not seem to satisfy Theorem 7: for a b.f. s' to be compatible with s it suffices to have *one* focal element intersecting the core $\mathcal{C}_s$, *not all* of them.

Definition 9. *The* compatible subspace $C(s)$ *associated to a belief function s is the collection of all the b.f.s with focal elements included into the core of s:* $C(s) \doteq \{s' : \mathcal{C}_{s'} \subset \mathcal{C}_s\}$.

From Theorem 7 it follows that

Corollary 4. $C(s) = Cl(\{P_A : A \subset \mathcal{C}_s\})$.

The compatible space $C(s)$ is only a *proper* subset of the collection of belief functions combinable with s, $\mathcal{S} \setminus NC(s)$: nevertheless, it contains all the relevant information. In fact,

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

Proof. Let us denote with $\mathcal{E}_{s'} = \{A_i\}$ and $\mathcal{E}_s = \{B_j\}$ the focal elements of s' and s respectively. Obviously $B_j \cap A_i = B_j \cap A_i \cap \mathcal{C}_s = B_j \cap (A_i \cap \mathcal{C}_s)$ so that defining a new b.f. s'' with focal elements

$$\mathcal{A}_i \doteq A_i \cap \mathcal{C}_s$$

and basic probability assignment $m''(\mathcal{A}_i) = m'(A_i)$ we have $s \oplus s' = s \oplus s''$. □

Now we are ready to formulate the geometric description of conditional subspaces. From Theorem 7 and 9 it comes directly

Corollary 5. $\langle s \rangle = Cl(\{s \oplus P_A, A \subset \mathcal{C}_s\})$.

Note that $s \oplus P_{\mathcal{C}_s} = s$, hence s is always a vertex of $\langle s \rangle$. Of course $\langle s \rangle \subset C(s)$, since the core is a *monotone function* on the poset $(\mathcal{S}, \geq_\oplus)$. Furthermore

$$dim(\langle s \rangle) = 2^{|\mathcal{C}_s|} - 2 \tag{7}$$

for the dimension of $\langle s \rangle$ is simply the cardinality of $C(s)$ (note that $\emptyset$ is not included) minus 1.
We can observe that

$$dim(NC(s)) + dim(\langle s \rangle) \neq dim(\mathcal{S}).$$

Corollary 5 depicts, in a sense, the *global* action of the orthogonal sum in the belief space. In [2] we started to analyze the *pointwise* behavior of Dempster's rule in $\mathcal{S}$, and its relation with the polytopes of probabilities consistent with the belief functions to combine.

8 Conclusions and perspectives

The geometric analysis exposed above is still at its initial stage, even if some interesting results have been achieved. We now have a picture of the behavior of belief functions as geometrical objects, but many questions still need to be addressed.
Some work has already been done on probabilistic ([12], [16]) and possibilistic ([5]) approximations of belief functions. For instance, F. Voorbraak proposed the following Bayesian approximation:

$$m(A) = \begin{cases} \frac{\sum_{B \supseteq A} m(B)}{\sum_{C \subseteq \Theta} m(C) \cdot |C|}, & |A| = 1 \\ 0 & otherwise \end{cases}$$

Nevertheless, we think that the geometric framework of the belief space could be the right context in which to pose and then solve the problem. Since a belief function is useful only when it is combined with others in an automated reasoning process, we can claim that *a good approximation, when combined with any other belief function, produces results similar to what obtained by combining the original function.* Analytically,

$$\hat{s} = \arg\min_{s' \in \mathcal{C}} \int_{t \in C(s)} dist(s \oplus t, s' \oplus t) dt \qquad (8)$$

where $dist$ is one of the classical L_p distance functions, and $\mathcal{C}$ is the class of belief functions where the approximation must belong. It can be proved (see [2] again) that for the simplest (binary) frame,

Proposition 4. *For every belief function $s \in \mathcal{S}_2$, the probabilistic approximation induced by the cost function (8) is unique, and corresponds to the normalized plausibility of singletons* for every arbitrary choice of the distance function L_p $\forall p$.

This suggests that the optimal approximation can be computed in closed form. Furthermore, the proposed criterion has a general scope, rests on intuitive principles and could be adopted to solve a wide number of problems.

On the other side, it is easy to see that, given the shape of conditional subspaces proved in Theorem 9, the simple components (e_1, e_2) of an arbitrary separable support function s in $\mathcal{S}_2$ can be expressed as

$$e_i = Cl(s, s \oplus P_i) \cap Cl(0, P_i) = Cl(s, P_i) \cap Cl(0, P_i).$$

Hence it seems likely that the language we introduced, based on the two operators of convex closure and orthogonal sum, could be powerful enough to provide a general solution to the canonical decomposition problem, alternative to Smets' ([13]) and Kramosil's ([8]) ones.

The lack of an evidential analogous of the notion of random process is perhaps one of the major drawbacks of the theory of evidence (as we mentioned in the Introduction) preventing a wider application to engineering problems. The knowledge of the geometrical form of conditional subspaces could be useful to predict the behavior of the *series of belief functions*

$$\lim_{n \to \infty} (s_1 \oplus \cdots \oplus s_n)$$

and their asymptotic properties.

In conclusion, we can argue that even if the geometrical analysis of the space of the belief functions was originally motivated by the approximation problem its potential applications are far more extended, and deserve further attentions.

References

[1] M. 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.

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

[3] F. Cuzzolin and R. Frezza, *An evidential reasoning framework for object tracking*, SPIE - Photonics East 99, vol. 3840, 19-22 September 1999, pp. 13–24.

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

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

[6] R. Fagin and Joseph Y. Halpern, *A new approach to updating beliefs*, Uncertainty in Artificial Intelligence, 6, 1991, pp. 347–374.

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

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

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

[10] J. D. Lowrance, T. D. Garvey, and T. M. Strat, *A framework for evidential-reasoning systems*, Proc. of the 5th National Conference on Artificial Intelligence, 1986, pp. 896–903.

[11] G. Shafer, *A mathematical theory of evidence*, Princeton University Press, 1976.

[12] Ph. Smets, *Belief functions versus probability functions*, Uncertainty and Intelligent Systems, Springer Verlag, Berlin, 1988, pp. 17–24.

[13] ———, *The canonical decomposition of a weighted belief*, Proceedings of IJCAI95, Montréal, Canada, 1995, pp. 1896–1901.

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

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

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

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