On the orthogonal projection of a belief function

Fabio Cuzzolin
Fabio.Cuzzolin@inrialpes.fr

INRIA Rhone-Alpes

Abstract. In this paper we study a new probability associated with any given belief function b, i.e. the orthogonal projection $\pi[b]$ of b onto the probability simplex $\mathcal{P}$. We provide an interpretation of $\pi[b]$ in terms of a redistribution process in which the mass of each focal element is equally distributed among its subsets, establishing an interesting analogy with the pignistic transformation. We prove that orthogonal projection commutes with convex combination just as the pignistic function does, unveiling a decomposition of $\pi[b]$ as convex combination of basis pignistic functions. Finally we discuss the norm of the difference between orthogonal projection and pignistic function in the case study of a quaternary frame, as a first step towards a more comprehensive picture of their relation.

1 Introduction

The theory of evidence (ToE) is one of the most popular uncertainty theories, thanks perhaps to its nature of quite natural extension of the classical Bayesian methodology. Indeed, the notion of *belief function* (b.f.) [1] generalizes that of finite probability, with classical probabilities forming a subclass $\mathcal{P}$ of b.f.s called *Bayesian* belief functions. The interplay of belief and Bayesian functions is of course of great interest in the theory of evidence. In particular, many people worked on the problem of finding a probabilistic approximation of an arbitrary belief function. Several papers [2–10] have been published on this issue, mainly in order to find efficient implementations of the rule of combination aiming to reduce the number of focal elements. The connection between belief functions and probabilities is as well crucial in Smets' "Transferable Belief Model" [11].
The study of the links between belief functions and probabilities has recently been posed in a geometric setup [12, 13]. In robust Bayesian statistics, there is a large literature on the study of convex sets of probability distributions [14–17]. On our side, in a series of works [18, 19] we proposed a geometric interpretation of the theory of evidence in which belief functions are represented as points of a simplex called *belief space* $\mathcal{B}$, a polytope whose vertices are all the b.f.s focused on a single event A, $m_b(A) = 1$, $m_b(B) = 0\ \forall B \neq A$. The region $\mathcal{P}$ of Bayesian b.f.s is also a simplex, part of the border of $\mathcal{B}$. The relation between belief and probability measures can then be naturally studied in this framework.
In this paper we use tools provided by the geometric approach to introduce a new probability function $\pi[b]$ associated with any given belief function b, precisely the

orthogonal projection of b onto the probability simplex $\mathcal{P}$. We thoroughly discuss its interpretation and properties, and its relations with other known Bayesian approximations of belief functions, i.e. pignistic function and relative plausibility of singletons. We show that $\pi[b]$ is inherently related to a redistribution process similar to that of the pignistic transformation, in which though the mass of each focal element is reassigned to *all its subsets*. We prove that, just as the pignistic function does, the orthogonal projection *commutes* with respect to the convex combination operator, yielding an interesting decomposition of $\pi[b]$ in terms of convex combination of basis pignistic functions.

2 A geometric approach to the ToE

A *basic belief assignment* (b.b.a.) over a finite set or "frame of discernment" [1] Θ is a function $m : 2^\Theta \rightarrow [0,1]$ on its power set $2^\Theta \doteq \{A \subseteq \Theta\}$ such that $m(\emptyset) = 0$, $\sum_{A \subseteq \Theta} m(A) = 1$, $m(A) \geq 0 \ \forall A \subseteq \Theta$. Subsets of Θ associated with non-zero values of m are called *focal elements*. The *belief function* $b : 2^\Theta \rightarrow [0,1]$ associated with a basic belief assignment m on Θ is defined as: $b(A) = \sum_{B \subseteq A} m(B)$. The unique b.b.a. m_b associated with a given b.f. b can be recovered by means of the Moebius inversion formula $m_b(A) = \sum_{B \subseteq A} (-1)^{|A-B|} b(B)$. In the ToE a probability function is simply a special belief function assigning non-zero masses to singletons only (*Bayesian* b.f.): $m_b(A) = 0$, $|A| > 1$. A dual mathematical representation of the evidence encoded by a b.f. b is the *plausibility function* (pl.f.) $pl_b : 2^\Theta \rightarrow [0,1]$, where the plausibility $pl_b(A)$ of an event A is given by $pl_b(A) \doteq 1 - b(A^c) = 1 - \sum_{B \subseteq A^c} m_b(B) = \sum_{B \cap A \neq \emptyset} m_b(B)$ where A^c denotes the complement of A in Θ.

Motivated by the search for a meaningful probabilistic approximation of belief functions we introduced the notion of *belief space* [19], as the space of all belief functions defined on a given frame of discernment Θ. A belief function $b : 2^\Theta \rightarrow [0,1]$ is completely specified by its $N-1$ belief values $\{b(A), A \subseteq \Theta, A \neq \emptyset\}$, $N \doteq 2^{|\Theta|}$, and can then be represented as a point of $\mathbb{R}^{N-1}$. The belief space associated with Θ is the set of points $\mathcal{B}_\Theta$ of $\mathbb{R}^{N-1}$ corresponding to a belief function. We will assume the domain Θ fixed, and denote the belief space with $\mathcal{B}$. It is not difficult to prove [19] that $\mathcal{B}$ is convex. Let $b_A \doteq b \in \mathcal{B}$ s.t. $m_b(A) = 1$, $m_b(B) = 0 \ \forall B \neq A$ be the unique belief function assigning all the mass to a single subset A of Θ (A-th *basis belief function*). It can be proved that [19] the belief space $\mathcal{B}$ is the convex closure of all basis belief functions b_A: $\mathcal{B} = Cl(b_A, \emptyset \subsetneq A \subseteq \Theta)$, where Cl denotes the convex closure operator:

$$Cl(b_1, ..., b_k) = \Big\{ b \in \mathcal{B} : b = \alpha_1 b_1 + \cdots + \alpha_k b_k : \textstyle\sum_i \alpha_i = 1, \alpha_i \geq 0 \ \forall i \Big\}. \quad (1)$$

The convex space delimited by a collection of (affinely independent [20]) points is called *simplex*. Each b.f. $b \in \mathcal{B}$ can be written as a convex sum as $b = \sum_{\emptyset \subsetneq A \subseteq \Theta} m_b(A) b_A$. Geometrically, the b.b.a. m_b is the set of coordinates of b in the simplex $\mathcal{B}$. Analogously, the set $\mathcal{P}$ of all Bayesian belief functions on

Θ is the simplex determined by all the basis b.f.s associated with singletons: $\mathcal{P} = Cl(b_x, x \in \Theta)$. PL.F.s can also be seen as points of $\mathbb{R}^{N-1}$ [19].

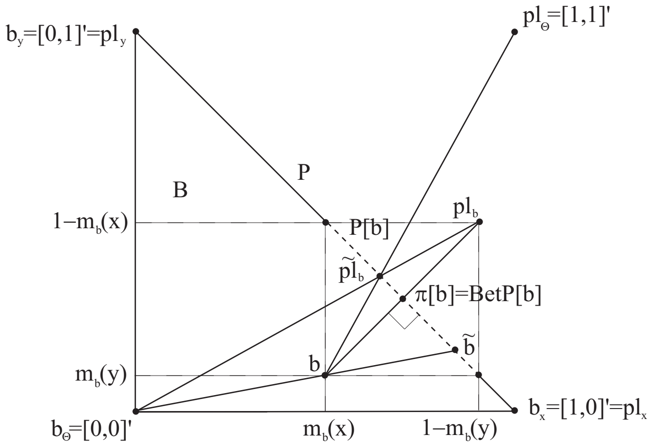

Fig. 1. In a binary frame $\Theta_2 = \{x, y\}$ the belief space $\mathcal{B}$ is a simplex with vertices $\{b_\Theta = [0,0]', b_x = [1,0]', b_y = [0,1]'\}$. A belief function b and the corresponding plausibility function pl_b are always located in symmetric positions with respect to the set $\mathcal{P}$ of probabilities on Θ. The associated relative plausibility $\tilde{pl}_b$ and belief $\tilde{b}$ of singletons are shown as the intersections of the probabilistic subspace with the line joining pl_b and $b_\Theta = [0,0]'$ and the line passing through b and b_Θ respectively. The other Bayesian functions related to b all coincide with the center of the segment of consistent probabilities $\mathcal{P}[b]$.

3 Orthogonal projection: binary case

It may be helpful to visually render these concepts in a simple example. Figure 1 shows the geometry of belief and plausibility functions for a binary frame $\Theta_2 = \{x, y\}$. Since $b(\Theta) = pl_b(\Theta) = 1$ for all b, we can represent belief and plausibility vectors as points of a plane with coordinates $b = [b(x) = m_b(x), b(y) = m_b(y)]'$, $pl_b = [pl_b(x) = 1 - m_b(y), pl_b(y) = 1 - m_b(x)]'$ respectively.
Each pair of functions (b, pl_b) determines a line which is *orthogonal* to $\mathcal{P}$, where b and pl_b lie on symmetric positions on the two sides of the Bayesian region. The set of probabilities $\mathcal{P}[b]$ consistent with b, i.e. $\mathcal{P}[b] \doteq \{p \in \mathcal{P} : p(A) \geq b(A) \ \forall A \subseteq \Theta\}$ in the simple binary case forms a segment in $\mathcal{P}$ (see Figure 1 again), whose center of mass is well known [21, 22, 18] to be Smets' *pignistic function* [23]

$$BetP[b] = \sum_{x \in \Theta} b_x \sum_{A \supseteq \{x\}} \frac{m_b(A)}{|A|} = b_x \Big(m_b(x) + \frac{m_b(\Theta)}{2} \Big) + b_y \Big(m_b(y) + \frac{m_b(\Theta)}{2} \Big). \tag{2}$$

It is interesting to notice though that it also coincides with the *orthogonal projection* $\pi[b]$ of b onto $\mathcal{P}$: $\pi[b] = BetP[b] = \overline{\mathcal{P}}[b]$. On the other side, both *relative plausibility* $\tilde{pl}_b$ and *relative belief* $\tilde{b}$ *of singletons*

$$\tilde{pl}_b(x) \doteq \frac{pl_b(x)}{\sum_{y\in\Theta} pl_b(y)}, \;\; \tilde{b}(x) \doteq \frac{b(x)}{\sum_{y\in\Theta} b(y)} \tag{3}$$

even though consistent with b, do not follow the same scheme.
In the following we will study the geometry of the orthogonal complement of $\mathcal{P}$ and analyze the properties of the associated Bayesian function, the orthogonal projection $\pi[b]$ of b onto the probability simplex $\mathcal{P}$.

4 Orthogonal projection

Let us then denote with $a(v_1, .., v_k)$ the affine subspace of some Cartesian space $\mathbb{R}^m$ generated by the points $v_1, ..., v_k \in \mathbb{R}^m$, i.e. the set $\{v \in \mathbb{R}^m : v = \alpha_1 v_1 + \cdots + \alpha_k v_k, \sum_i \alpha_i = 1\}$. The orthogonal projection $\pi[b]$ of b onto $a(\mathcal{P})$ is obviously guaranteed to exist as $a(\mathcal{P})$ is nothing but a linear subspace of $\mathbb{R}^{N-1}$ (on which b lives). By definition, the orthogonal projection $\pi[b]$ is the solution of the optimization problem

$$\pi[b] = \arg\min_{p\in\mathcal{P}} \|p - b\|_{L_2} = \arg\min_{p\in\mathcal{P}} \sqrt{\sum_{A\subseteq\Theta} |p(A) - b(A)|^2} \tag{4}$$

and is then naturally the unique solution of the Bayesian approximation problem when choosing the L_2 distance in the belief space.
An explicit calculation of $\pi[b]$ requires a description of the orthogonal complement of $a(\mathcal{P})$ in $\mathbb{R}^{N-1}$. Let $n = |\Theta|$ be the cardinality of Θ.

4.1 General form of the orthogonal projection

We need to find a necessary and sufficient condition for an arbitrary vector $v = \sum_{A\subseteq\Theta} v_A X_A$, where $\{X_A, A \subseteq \Theta\}$ is a reference frame in $\mathbb{R}^{N-1}$, to be orthogonal to the probabilistic subspace $a(\mathcal{P})$. If we compute the scalar product $\langle v, b_y - b_x\rangle$ between v and the generators $b_y - b_x$ of $a(P)$ we get

$$\Big\langle \sum_{A\subseteq\Theta} v_A X_A, b_y - b_x \Big\rangle = \sum_{A\subseteq\Theta} v_A [b_y - b_x](A).$$

After remembering that, by definition, $b_A(B) = 1$ if $B \supseteq A$, 0 elsewhere, we can see that these vectors display a special symmetry

$$b_y - b_x(A) = \begin{cases} 1 & A \supseteq \{y\}, A \not\supset \{x\} \\ 0 & A \supseteq \{x\}, \{y\} \; or \; A \not\supset \{x\}, \{y\} \\ -1 & A \not\supset \{y\}, A \supseteq \{x\} \end{cases}$$

which allows us to write $\langle v, b_y - b_x \rangle = \sum_{A \supseteq y, A \not\supset \{x\}} v_A - \sum_{A \supseteq \{x\}, A \not\supset \{y\}} v_A$. The orthogonal complement $a(\mathcal{P})^\perp$ of $a(\mathcal{P})$ will then be expressed as

$$a(\mathcal{P})^\perp = \Big\{ v : \sum_{A \supseteq \{y\}, A \not\supset \{x\}} v_A = \sum_{A \supseteq \{x\}, A \not\supset \{y\}} v_A \;\forall y \neq x \Big\}.$$

If the vector v, in particular, is a belief function

$$\sum_{A \supseteq \{y\}, A \not\supset \{x\}} b(A) = \sum_{A \supseteq \{y\}, A \not\supset \{x\}} \sum_{B \subseteq A} m_b(B) = \sum_{B \subseteq \{x\}^c} m_b(B) \cdot 2^{n-1-|B \cup \{y\}|}$$

since $2^{n-1-|B \cup \{y\}|}$ is the number of subsets A of $\{x\}^c$ containing both B and y, and the orthogonality condition becomes

$$\sum_{B \subseteq \{x\}^c} m_b(B) 2^{1-|B \cup \{y\}|} = \sum_{B \subseteq \{y\}^c} m_b(B) 2^{1-|B \cup \{x\}|}$$

$\forall y \neq x$, after erasing the common factor 2^{n-2}. Now, events $B \subseteq \{x, y\}^c$ appear in both summations, with the same coefficient (since $|B \cup \{x\}| = |B \cup \{y\}| = |B| + 1$) and the equation reduces to

$$\sum_{B \supseteq \{y\}, B \not\supset \{x\}} m_b(B) 2^{1-|B|} = \sum_{B \supseteq \{x\}, B \not\supset \{y\}} m_b(B) 2^{1-|B|} \tag{5}$$

$\forall y \neq x$, the desired orthogonality condition. (5) can be used to prove that [24]

Theorem 1. *The orthogonal projection $\pi[b]$ of b onto $a(\mathcal{P})$ can be expressed in terms of the b.b.a. m_b of b in the following equivalent forms:*

$$\sum_{A \supseteq \{x\}} m_b(A) 2^{1-|A|} + \sum_{A \subseteq \Theta} m_b(A) \Big(\frac{1 - |A| 2^{1-|A|}}{n} \Big) \tag{6}$$

$$\sum_{A \supseteq \{x\}} m_b(A) \Big(\frac{1 + |A^c| 2^{1-|A|}}{n} \Big) + \sum_{A \not\supset \{x\}} m_b(A) \Big(\frac{1 - |A| 2^{1-|A|}}{n} \Big). \tag{7}$$

From (7) we can see that $\pi[b]$ *is indeed a probability*, since both $1 + |A^c| 2^{1-|A|} \geq 0$ and $1 - |A| 2^{1-|A|} \geq 0$ $\forall |A| = 1, ..., n$. This is not at all trivial, as $\pi[b]$ is the projection of b onto the *affine* space $a(\mathcal{P})$, and could have in principle assigned negative masses to one or more singletons. $\pi[b]$ is hence a valid candidate to the role of probabilistic approximation of the b.f. b.

Unnormalized case It is interesting to note that the above results hold for *unnormalized* belief functions [25] too. The orthogonality results of Section 4.1 are still valid as the proof of Theorem 1 [24] does not concern the mass of the empty set. The orthogonal projection $\pi[b]$ of a u.b.f. b is then well defined and is still given by Equations (6),(7) where this time the summations on the right hand side include $\emptyset$ too:

$$\pi[b](x) = \sum_{A\supseteq\{x\}} m_b(A)2^{1-|A|} + \sum_{\emptyset\subseteq A\subseteq\Theta} m_b(A)\Big(\frac{1-|A|2^{1-|A|}}{n}\Big)$$

$$\pi[b](x) = \sum_{A\supseteq\{x\}} m_b(A)\Big(\frac{1+|A^c|2^{1-|A|}}{n}\Big) + \sum_{\emptyset\subseteq A\not\supset\{x\}} m_b(A)\Big(\frac{1-|A|2^{1-|A|}}{n}\Big).$$

4.2 Orthogonality flag and redistribution process

Theorem 1 does not apparently provide any intuition about the meaning of $\pi[b]$ in terms of degrees of belief. If we process Equation (7) though we can reduce π to a new Bayesian function strictly related to the pignistic function [24].

Theorem 2. *$\pi[b] = \bar{\mathcal{P}}(1-k_O[b])+k_O[b]O[b]$, where $\bar{\mathcal{P}}$ is the uniform probability and $O[b](x)$ is the Bayesian b.f.*

$$O[b](x) = \frac{\bar{O}[b](x)}{k_O[b]} = \frac{\sum_{A\supseteq\{x\}} m_b(A)2^{1-|A|}}{\sum_{A\subseteq\Theta} m_b(A)|A|2^{1-|A|}}. \tag{8}$$

As $0 \leq |A|2^{1-|A|} \leq 1$ for all $A \subseteq \Theta$, $k_O[b]$ assumes values in the interval $[0,1]$. Theorem 2 then implies that the orthogonal projection is always located on the line segment joining the uniform, non-informative probability $\bar{\mathcal{P}}$ and the Bayesian b.f. $O[b]$. By Equation (8) it turns out that $\pi[b] = \bar{\mathcal{P}}$ iff $O[b] = \bar{\mathcal{P}}$ (since $k_O[b] > 0$). The meaning to attribute to $O[b]$ becomes clear when we notice that the condition (5) under which a b.f. b is orthogonal to $a(\mathcal{P})$ can be rewritten as $\sum_{B\supseteq\{y\},B\not\supset\{x\}} m_b(B)2^{1-|B|} + \sum_{B\supseteq\{y\},\{x\}} m_b(B)2^{1-|B|} = \sum_{B\supseteq\{x\},B\not\supset\{y\}} m_b(B)2^{1-|B|}+\sum_{B\supseteq\{y\},\{x\}} m_b(B)2^{1-|B|} \equiv \sum_{B\supseteq\{y\}} m_b(B)2^{1-|B|} = \sum_{B\supseteq\{x\}} m_b(B)2^{1-|B|} \equiv \bar{O}[b](x) = const \equiv O[b](x) = const = \bar{\mathcal{P}}\ \forall\ x \in \Theta$. Therefore $\pi[b] = \bar{\mathcal{P}}$ iff $b\perp a(\mathcal{P})$, and $O - \bar{\mathcal{P}}$ measures the non-orthogonality of b with respect to $\mathcal{P}$. $O[b]$ deserves then the name of *orthogonality flag*.

A compelling link can be drawn between orthogonal projection and pignistic function through the orthogonality flag $O[b]$. Let us define the two b.f.

$$b_{||} \doteq \frac{1}{k_{||}} \sum_{A\subseteq\Theta} \frac{m_b(A)}{|A|} b_A, \qquad b_{2||} \doteq \frac{1}{k_{2||}} \sum_{A\subseteq\Theta} \frac{m_b(A)}{2^{|A|}} b_A,$$

$k_{||}$ and $k_{2||}$ their normalization factors.

Theorem 3. *$O[b]$ is the relative plausibility of singletons of $b_{2||}$; $BetP[b]$ is the relative plausibility of singletons of $b_{||}$.*

Proof. By definition of plausibility function

$$pl_{b_{2||}}(x) = \sum_{A\supseteq\{x\}} m_{b_{2||}}(A) = \frac{1}{k_{2||}} \sum_{A\supseteq\{x\}} \frac{m_b(A)}{2^{|A|}} = \frac{\bar{O}[b]}{2k_{2||}},$$

$$\sum_{x\in\Theta} pl_{b_{2||}}(x) = \frac{1}{k_{2||}} \sum_{x\in\Theta} \sum_{A\supseteq\{x\}} \frac{m_b(A)}{2^{|A|}} = \frac{k_O[b]}{2k_{2||}}$$

as $k_O[b]$ is the normalization factor for $\bar{O}[b]$:

$$\sum_{x\in\Theta}\bar{O}[b](x)=\sum_{x\in\Theta}\sum_{A\supseteq\{x\}}m_b(A)2^{1-|A|}=\sum_{A\subseteq\Theta}m_b(A)|A|2^{1-|A|}=k_O[b].$$

Hence $\tilde{pl}_{b_{2||}}(x)=\bar{O}[b]/k_O[b]=O[b]$, i.e.

$$pl_{b_{||}}(x)=\sum_{A\supseteq\{x\}}m_{b_{||}}(A)=\frac{1}{k_{||}}\sum_{A\supseteq\{x\}}\frac{m_b(A)}{|A|}=\frac{1}{k_{||}}BetP[b](x)$$

and since $\sum_x BetP[b](x)=1$, $\tilde{pl}_{b_{||}}(x)=BetP[b](x)$.

The two functions $b_{||}$ and $b_{2||}$ represent two different processes acting on b (see Figure 2). The first one redistributes the mass of each focal element among its *singletons* (yielding directly a Bayesian b.f. $BetP[b]$). The second one distributes the b.b.a. of each event A among its *subsets* $B\subseteq A$ ($\emptyset, A$ included). In this second case we get an *unnormalized* [25] b.f. b^U with $m_{b^U}(A)=\sum_{B\supseteq A}\frac{m_b(B)}{2^{|B|}}$ whose relative belief of singletons (3) $\tilde{b^U}$ is in fact the orthogonality flag $O[b]$.

Example Let us consider as an example the belief function b on the ternary frame: $m_b(x)=0.1$, $m_b(y)=0$, $m_b(z)=0.2$, $m_b(\{x,y\})=0.3$, $m_b(\{x,z\})=0.1$, $m_b(\{y,z\})=0$, $m_b(\Theta)=0.3$. To get the orthogonality flag $O[b]$ we need to apply the redistribution process of Figure 2 to each focal element of b. In this case their masses are divided among their subsets as follows:

$$\begin{array}{ll}
m_b(x)=0.1 & \mapsto m'(x)=m'(\emptyset)=0.1/2=0.05,\\
m_b(z)=0.2 & \mapsto m'(z)=m'(\emptyset)=0.2/2=0.1,\\
m_b(\{x,y\})=0.3 & \mapsto m'(\{x,y\})=m'(x)=m'(y)=m'(\emptyset)=0.3/4=0.075,\\
m_b(\{x,z\})=0.1 & \mapsto m'(\{x,z\})=m'(x)=m'(z)=m'(\emptyset)=0.1/4=0.025,\\
m_b(\Theta)=0.3 & \mapsto m'(\Theta)=m'(\{x,y\})=m'(\{x,z\})=m'(\{y,z\})=\\
 & \quad =m'(x)=m'(y)=m'(z)=m'(\emptyset)=0.3/8=0.0375.
\end{array}$$

By summing contributions related to singletons on the right hand side we get

$$\begin{array}{c}
m_{b^U}(x)=0.05+0.075+0.025+0.0375=0.1875,\\
m_{b^U}(y)=0.075+0.0375=0.1125, \quad m_{b^U}(z)=0.1+0.025+0.0375=0.1625
\end{array}$$

whose sum is the normalization factor $k_O[b]=m_{b^U}(x)+m_{b^U}(y)+m_{b^U}(z)=0.4625$ and by normalizing $O[b]=[0.405,0.243,0.351]'$. The orthogonal projection $\pi[b]$ is finally the convex combination of $O[b]$ and $\bar{\mathcal{P}}=[1/3,1/3,1/3]'$ with coordinate $k_O[b]$: $\pi[b]=\bar{\mathcal{P}}(1-k_O[b])+k_O[b]O[b]=[1/3,1/3,1/3]'\cdot(1-0.4625)+0.4625\cdot[0.405,0.243,0.351]'=[0.366,0.291,0.342]'$.

4.3 Orthogonal projection and convex combination

As a confirmation of this relationship, orthogonal projection and pignistic function both commute with convex combination.

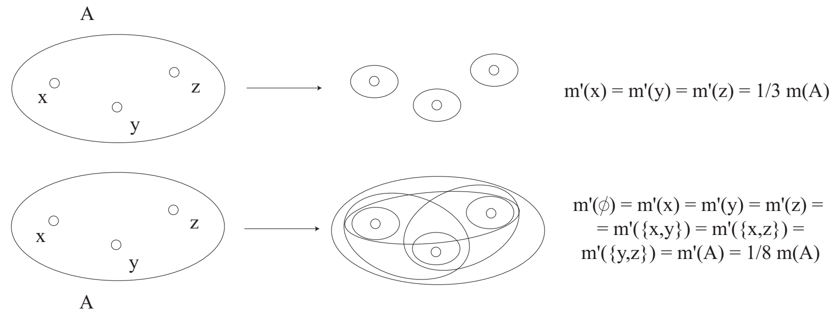

Fig. 2. Redistribution processes associated with pignistic transformation and orthogonal projection. In the pignistic transformation (top) the mass of each focal element is distributed among its elements. In the orthogonal projection (bottom), instead (through the orthogonality flag), the mass of each f.e. is divided among its subsets. In both cases, the related relative belief of singletons yields a Bayesian belief function.

Theorem 4. *Orthogonal projection and convex combination commute, i.e. if* $\alpha_1 + \alpha_2 = 1$ *then* $\pi[\alpha_1 b_1 + \alpha_2 b_2] = \alpha_1 \pi[b_1] + \alpha_2 \pi[b_2]$.

Proof. By Theorem 2 $\pi[b] = (1 - k_O[b])\bar{\mathcal{P}} + \bar{O}[b]$ where the coefficient is $k_O[b] = \sum_{A \subseteq \Theta} m_b(A)|A|2^{1-|A|}$ and $\bar{O}[b](x) = \sum_{A \supseteq \{x\}} m_b(A) 2^{1-|A|}$. Hence

$$k_O[\alpha_1 b_1 + \alpha_2 b_2] = \sum_{A \subseteq \Theta} \big(\alpha_1 m_{b_1}(A) + \alpha_2 m_{b_2}(A)\big)|A|2^{1-|A|} = \alpha_1 k_O[b_1] + \alpha_2 k_O[b_2],$$
$$\bar{O}[\alpha_1 b_1 + \alpha_2 b_2](x) = \sum_{A \supseteq \{x\}} \big(\alpha_1 m_{b_1}(A) + \alpha_2 m_{b_2}(A)\big)2^{1-|A|} = \alpha_1 \bar{O}[b_1] + \alpha_2 \bar{O}[b_2]$$

which in turn implies (since $\alpha_1 + \alpha_2 = 1$) $\pi[\alpha_1 b_1 + \alpha_2 b_2] = (1 - \alpha_1 k_O[b_1] - \alpha_2 k_O[b_2])\bar{\mathcal{P}} + \alpha_1 \bar{O}[b_1] + \alpha_2 \bar{O}[b_2] = \alpha_1 \big[(1 - k_O[b_1])\bar{\mathcal{P}} + \bar{O}[b_1]\big] + \alpha_2 \big[(1 - k_O[b_2])\bar{\mathcal{P}} + \bar{O}[b_2]\big] = \alpha_1 \pi[b_1] + \alpha_2 \pi[b_2]$.

This property can be used to find an alternative expression of the orthogonal projection as convex combination of the pignistic functions associated with all basis belief functions.

Lemma 1. *The orthogonal projection of a basis belief function* b_A *is given by*

$$\pi[b_A] = (1 - |A|2^{1-|A|})\bar{\mathcal{P}} + |A|2^{1-|A|}\bar{\mathcal{P}}_A,$$

with $\bar{\mathcal{P}}_A = \frac{1}{|A|}\sum_{x \in A} b_x$ *the center of mass of all probabilities with support in* A.

Proof. By Equation (8) $k_O[b_A] = |A|2^{1-|A|}$, so that $\bar{O}[b_A](x) = 2^{1-|A|}$ if $x \in A$, 0 otherwise. This implies

$$O[b_A](x) = \begin{cases} \frac{1}{|A|} & x \in A \\ 0 & x \notin A \end{cases} = \frac{1}{|A|}\sum_{x \in A} b_x = \bar{\mathcal{P}}_A.$$

Theorem 5. *The orthogonal projection can be expressed as a convex combination of all the non-informative probabilities with support on a single event A:*

$$\pi[b] = \bar{\mathcal{P}}\Big(1 - \sum_{A \neq \Theta} \alpha_A\Big) + \sum_{A \neq \Theta} \alpha_A \bar{\mathcal{P}}_A, \qquad \alpha_A \doteq m_b(A)|A|2^{1-|A|}. \tag{9}$$

Proof. $\pi[b] = \pi\Big[\sum_{A\subseteq\Theta} m_b(A) b_A\Big] = \sum_{A\subseteq\Theta} m_b(A)\pi[b_A]$ by Theorem 4, which by Lemma 1 becomes $\sum_{A\subseteq\Theta} m_b(A)[(1-|A|2^{1-|A|})\bar{\mathcal{P}} + |A|2^{1-|A|}\bar{\mathcal{P}}_A] = (1 - \sum_{A\subseteq\Theta} m_b(A)|A|2^{1-|A|})\bar{\mathcal{P}} + \sum_{A\subseteq\Theta} m_b(A)|A|2^{1-|A|}\bar{\mathcal{P}}_A$ which can be written as

$$\Big(1 - \sum_{A\subseteq\Theta} m_b(A)|A|2^{1-|A|}\Big)\bar{\mathcal{P}} + \sum_{A\neq\Theta} m_b(A)|A|2^{1-|A|}\bar{\mathcal{P}}_A + m_b(\Theta)|\Theta|2^{1-|\Theta|}\bar{\mathcal{P}}$$

i.e. Equation (9).

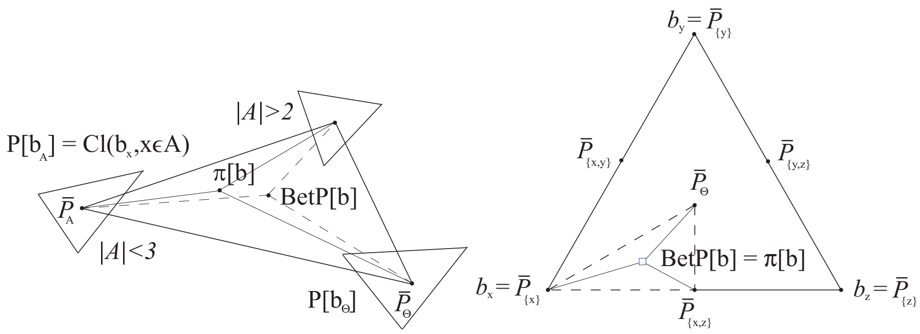

Fig. 3. Left: Orthogonal projection $\pi[b]$ and pignistic function $BetP[b]$ are both located on the simplex whose vertices are all the basis pignistic functions, i.e. the uniform probabilities associated with each single event A. However, the convex coordinates of $\pi[b]$ are weighted by a factor $k_O[b_A] = |A|2^{1-|A|}$, yielding a point which is closer to vertices related to lower size events. Right: Orthogonal projection and pignistic function coincide in the ternary case $\Theta_3 = \{x, y, z\}$.

As $\bar{\mathcal{P}}_A = BetP[b_A]$, we recognize that

$$BetP[b] = \sum_{A\subseteq\Theta} m_b(A) BetP[b_A],$$
$$\pi[b] = \sum_{A\neq\Theta} \alpha_A BetP[b_A] + \Big(1 - \sum_{A\neq\Theta}\alpha_A\Big) BetP[b_\Theta],$$

i.e. both orthogonal projection and pignistic function are convex combinations of all the basis pignistic functions. However, as $k_O[b_A] = |A|2^{1-|A|} < 1$ for $|A| > 2$, the orthogonal projection turns out to be closer to the vertices associated with events of lower cardinality (see Figure 3-left).

Example: ternary case Let us consider as an example a ternary frame $\Theta_3 = \{x, y, x\}$, and a belief function on Θ_3 with b.b.a. $m_b(x) = 1/3$, $m_b(\{x, z\}) = 1/3$, $m_b(\Theta_3) = 1/3$, $m_b(A) = 0$ $A \neq \{x\}, \{x, z\}, \Theta_3$. According to Equation (9)

$$\begin{gathered}\pi[b] = 1/3\bar{\mathcal{P}}_{\{x\}} + 1/3\bar{\mathcal{P}}_{\{x,z\}} + (1 - 1/3 - 1/3)\bar{\mathcal{P}} = \tfrac{1}{3}b_x + \tfrac{1}{3}\tfrac{b_x+b_z}{2} + \\ + \tfrac{1}{3}\tfrac{b_x+b_y+b_z}{3} = b_x(\tfrac{1}{3} + \tfrac{1}{6} + \tfrac{1}{9}) + b_z(\tfrac{1}{6} + \tfrac{1}{9}) + b_y\tfrac{1}{9} = \tfrac{11}{18}b_x + \tfrac{1}{9}b_y + \tfrac{5}{18}b_z\end{gathered}$$

and the orthogonal projection is the barycenter of the simplex $Cl(\bar{\mathcal{P}}_{\{x\}}, \bar{\mathcal{P}}_{\{x,z\}}, \bar{\mathcal{P}})$ (see Figure 3-right). On the other side $BetP[b](x) = m_b(x) + \frac{m_b(x,z)}{2} + \frac{m_b(\Theta_3)}{3} = \frac{11}{18}$, $BetP[b](y) = \frac{1}{9}$, $BetP[b](z) = \frac{1}{6} + \frac{1}{9} = \frac{5}{18}$ i.e. $BetP[b] = \pi[b]$. This is true for each belief function $b \in \mathcal{B}_3$, since for the above expressions when $|\Theta| = 3$ $\alpha_A = m_b(A)$ for $|A| \leq 2$, and $1 - \sum_A \alpha_A = 1 - \sum_{A\neq\Theta} m_b(A) = m_b(\Theta)$.

4.4 A quantitative analysis of the distance between $BetP$ and π

An exhaustive description of the relationship between orthogonal projection and pignistic function would require a quantitative analysis of their distance as the degrees of belief of b vary in the belief space.
Considered the fact that $\pi[b]$ is the solution of the Bayesian approximation problem when using the L_2 norm (4), a sensible choice is measuring their distance by computing the L_2 norm of their difference vector:

$$\|\pi[b] - BetP[b]\|_{L_2} \doteq \sqrt{\sum_{x\in\Theta} |\pi[b](x) - BetP[b](x)|^2}.$$

Let us then measure their difference in the simplest case in which they are distinct: a frame $\Theta = \{x, y, z, w\}$ of size 4. Their analytic expressions

$$\begin{aligned} BetP[b](x) = {} & \frac{1}{4}m_b(\Theta) + m_b(x) + \frac{1}{2}(m_b(\{x,y\}) + m_b(\{x,z\}) + m_b(\{x,w\})) + \\ & + \frac{1}{3}(m_b(\{x,y,z\}) + m_b(\{x,y,w\}) + m_b(\{x,z,w\})); \\ \pi[b](x) = {} & \frac{1}{4}m_b(\Theta) + m_b(x) + \frac{1}{2}(m_b(\{x,y\}) + m_b(\{x,z\}) + m_b(\{x,w\})) + \\ & + \frac{5}{16}(m_b(\{x,y,z\}) + m_b(\{x,y,w\}) + m_b(\{x,z,w\})) + \frac{1}{16}m_b(\{y,z,w\}) \end{aligned} \tag{10}$$

are very similar to each other. Basically the difference is that $\pi[b]$ counts also the masses of focal elements in $\{x\}^c$ (with a small contribution), while $BetP[b]$ by definition does not. If we compute their difference $BetP[b](x) - \pi[b](x) = \frac{1}{48}[m_b(\{x,y,z\}) + m_b(\{x,y,w\}) + m_b(\{x,z,w\}) - 3\, m_b(\{y,z,w\})]$ we can analyze the behavior of their L_2 distance as b varies. After introducing the simpler notation

$$y_1 = m_b(\{x,y,z\}),\ y_2 = m_b(\{x,y,w\}),\ y_3 = m_b(\{x,z,w\}),\ y_4 = m_b(\{y,z,w\}),$$

we can maximize (minimize) the norm above $(y_1 + y_2 + y_3 - 3y_4)^2 + (y_1 + y_2 + y_4 - 3y_3)^2 + (y_1 + y_3 + y_4 - 3y_2)^2 + (y_2 + y_3 + y_4 - 3y_1)^2$ by imposing

$\frac{\partial}{\partial y_i}\|BetP[b](\mathbf{y}) - \pi[b](\mathbf{y})\|^2 = 0$ subject to $y_1 + y_2 + y_3 + y_4 = 1$. The unique solution turns out to be $\mathbf{y} = [1/4, 1/4, 1/4, 1/4]'$ which corresponds to (after replacing this solution in (10) $BetP[b] = \pi[b] = \bar{\mathcal{P}}$ where $\bar{\mathcal{P}} = [1/4, 1/4, 1/4, 1/4]'$ is the uniform probability on Θ. The distance between pignistic function and orthogonal projection is minimal (zero) when all size 3 subsets have the same mass.
It is then natural to suppose that their difference must be maximal when all the mass is concentrated on a single size-3 event. This is in fact correct: $\|BetP[b] - \pi[b]\|^2$ is maximal and equal to $1^2 + 1^2 + 1^2 + (-3)^2 = 12$ when $y_i = 1$, $y_j = 0$ $\forall j \neq i$, i.e. the mass of one among $\{x, y, z\}, \{x, y, w\}, \{x, z, w\}, \{y, z, w\}$ is one.
Other distances could of course be chosen to assess the difference between Bayesian approximations in the probability simplex: A natural generalization of L_2 is the Mahalanobis distance $\sqrt{(p - p')'\Sigma(p - p')}$ (where Σ is a covariance matrix) which is often used in statistics. Our intuition on the problem suggests that the above results should hold for a wide class of such functions: Experimental validation is though needed.

5 Conclusions

In this paper we introduced a new Bayesian b.f. associated with any given belief function b, i.e. the orthogonal projection of b onto the probability simplex $\mathcal{P}$, by definition the solution of the probabilistic approximation problem when using the classical L_2 distance. Even though $\pi[b]$ has been derived through purely geometric considerations, it exhibits strong links with the pignistic function. Its interpretation in terms of rationality principles similar to those formulated for the pignistic transformation is still unclear, as it is to decide whether or not $\pi[b]$ is consistent with b. The redistribution process of Section 4.2 is a first step in this direction: The orthogonal projection is the result of a more "cautious" approach (with respect to $BetP$) in which the mass of higher-size events is not divided among singletons, but among subsets.

References

1. Shafer, G.: A Mathematical Theory of Evidence. Princeton University Press (1976)
2. Yaghlane, A.B., Denoeux, T., Mellouli, K.: Coarsening approximations of belief functions. In Benferhat, S., Besnard, P., eds.: Proceedings of ECSQARU'2001. (2001) 362–373
3. Cobb, B., Shenoy, P.: On the plausibility transformation method for translating belief function models to probability models. International Journal of Approximate Reasoning **41**(3) (April 2006) 314–330
4. Denoeux, T.: 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
5. Denoeux, T., Yaghlane, A.B.: 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

6. Haenni, R., Lehmann, N.: Resource bounded and anytime approximation of belief function computations. International Journal of Approximate Reasoning **31(1-2)** (October 2002) 103–154
7. Bauer, M.: Approximation algorithms and decision making in the Dempster-Shafer theory of evidence–an empirical study. International Journal of Approximate Reasoning **17** (1997) 217–237
8. Tessem, B.: Approximations for efficient computation in the theory of evidence. Artificial Intelligence **61:2** (1993) 315–329
9. Lowrance, J.D., Garvey, T.D., Strat, T.M.: A framework for evidential-reasoning systems. In for Artificial Intelligence, A.A., ed.: Proceedings of the National Conference on Artificial Intelligence. (1986) 896–903
10. Voorbraak, F.: A computationally efficient approximation of Dempster-Shafer theory. International Journal on Man-Machine Studies **30** (1989) 525–536
11. Smets, P.: Belief functions versus probability functions. In Bouchon B., S.L., R., Y., eds.: Uncertainty and Intelligent Systems. Springer Verlag, Berlin (1988) 17–24
12. Ha, V., Haddawy, P.: Geometric foundations for interval-based probabilities. In Cohn, A.G., Schubert, L., Shapiro, S.C., eds.: KR'98: Principles of Knowledge Representation and Reasoning. Morgan Kaufmann, San Francisco, California (1998) 582–593
13. Black, P.: An examination of belief functions and other monotone capacities. PhD dissertation, Department of Statistics, CMU (1996) Pgh. PA 15213.
14. Cozman, F.G.: Calculation of posterior bounds given convex sets of prior probability measures and likelihood functions. Journal of Computational and Graphical Statistics **8(4)** (1999) 824–838
15. Berger: Robust bayesian analysis: Sensitivity to the prior. Journal of Statistical Planning and Inference **25** (1990) 303–328
16. Herron, T., Seidenfeld, T., Wasserman, L.: Divisive conditioning: further results on dilation. Philosophy of Science **64** (1997) 411–444
17. Seidenfeld, T., Wasserman, L.: Dilation for convex sets of probabilities. Annals of Statistics **21** (1993) 1139–1154
18. Cuzzolin, F.: Geometry of upper probabilities. In: Proceedings of the 3^{rd} International Symposium on Imprecise Probabilities and Their Applications (ISIPTA'03). (July 2003)
19. Cuzzolin, F.: A geometric approach to the theory of evidence. IEEE Transactions on Systems, Man and Cybernetics part C (to appear) (2007)
20. Dubrovin, B., Novikov, S., Fomenko, A.: Geometria Contemporanea 3. Editori Riuniti (1989)
21. Chateauneuf, A., Jaffray, J.Y.: Some characterizations of lower probabilities and other monotone capacities through the use of Möbius inversion. Mathematical Social Sciences **17** (1989) 263–283
22. Dubois, D., Prade, H., Smets, P.: A definition of subjective possibility. International Journal of Approximate Reasoning (in press, March 2007)
23. Smets, P.: Decision making in the TBM: the necessity of the pignistic transformation. International Journal of Approximate Reasoning **38**(2) (February 2005) 133–147
24. Cuzzolin, F.: Two new Bayesian approximations of belief functions based on convex geometry. IEEE Transactions on Systems, Man and Cybernetics part B (2007)
25. Smets, P.: The nature of the unnormalized beliefs encountered in the Transferable Belief Model. In: Proceedings of UAI'92, San Mateo, CA, Morgan Kaufmann (1992) 292–29