Consistent approximations of belief functions

Fabio Cuzzolin
Oxford Brookes University, Oxford, UK
fabio.cuzzolin@brookes.ac.uk

Abstract

Consistent belief functions represent collections of coherent or non-contradictory pieces of evidence. As most operators used to update or elicit evidence do not preserve consistency, the use of consistent transformations $cs[\cdot]$ in a reasoning process to guarantee coherence can be desirable. Such transformations are turn linked to the problem of approximating an arbitrary belief function with a consistent one.

We study here the consistent approximation problem in the case in which distances are measured using classical L_p norms. We show that, for each choice of the element we want them to focus on, the partial approximations determined by the L_1 and L_2 norms coincide, and can be interpreted as classical focused consistent transformations. Global L_1 and L_2 solutions do not in general coincide, however, nor are they associated with the highest plausibility element.

Keywords. Consistent belief function, simplicial complex, approximation, L_p norms.

1 The consistent approximation problem

Belief functions (b.f.s) [19] are complex objects, in which different and sometimes contradictory bodies of evidence may coexist, as they mathematically describe the fusion of possibly conflicting expert opinions and/or imprecise/ corrupted measurements, etcetera. Making decisions based on such objects can then be misleading. This is a well known problem in classical logics, where the application of inference rules to inconsistent sets of assumptions or "knowledge bases" may lead to incompatible conclusions, depending on the subset of assumptions we start our reasoning from.

Consistent belief functions (cs.b.f.s), i.e. belief functions whose non-zero mass events or "focal elements" have non-empty intersection or "core", are then particularly interesting as they represent collections of coherent or non-contradictory pieces of evidence. In some situations it may then be desirable to design a method which, given an arbitrary belief function b, generates a consistent or non-contradictory belief function $cs[b]$: we call this *consistent transformation*. Such a transformation is all the more valuable as several important operators used to update or elicit evidence represented as belief measures, like Dempster's sum [8] and disjunctive combination [21], do not preserve consistency. To guarantee the consistency of a state of belief we may want to seek a scheme in which each time new evidence is combined to yield a new b.f., the consistent transformation $cs[\cdot]$ is applied to reduce it to a coherent knowledge state.

Now, consistent transformations can be built by solving a minimization problem of the form $cs[b] = \arg\min_{cs \in \mathcal{CS}} dist(b, cs)$, where $dist$ is some distance measure between belief functions, and $\mathcal{CS}$ denotes the collection of all consistent b.f.s. We call this the *consistent approximation problem*. By plugging in different distance functions we get different consistent transformations.

In this paper, in particular, we study what happens when using classical L_p norms. Indeed, consistent belief functions correspond to possibility distributions (Section 2), which are in turn inherently related to the L_∞ norm. Besides, the region of all cs.b.f.s is geometrically the set of belief functions for which the L_∞ norm of the plausibility distribution is equal to 1. We can then conjecture that L_p consistent approximations will be meaningful in terms of degrees of belief. This is indeed the case.

From a technical point of view, consistent b.f.s do not live in a single linear space, but in a collection of higher-dimensional triangles or simplices, called "simplicial complex" [11]. A partial solution has then to be found separately for each maximal simplex $\mathcal{CS}_x$ of the consistent complex $\mathcal{CS}$, i.e., the set of cs.b.f.s whose core includes the element x. These partial solutions are later to be compared to determine the global optimal solution.

We will prove here that the partial approximations determined by both the L_1 and the L_2 norms are unique and coincide. We will also prove that the L_1/L_2 consistent approximation onto each component $\mathcal{CS}_x$ of $\mathcal{CS}$ generates indeed the *consistent transformation focused on* x [10, 1], i.e. a new belief function whose focal elements have the form $A' = A \cup \{x\}$, where A is a focal element of the original b.f. b. As we will see, though, the associated global L_1/L_2 solutions do not lie in general on the same component of the consistent complex.

1.1 Paper outline

After recalling the notions of consistent and consonant belief functions, we will recall their semantics and stress why it can be desirable to transform a generic belief function into a consistent one (Section 2). As we pose the approximation problem in a geometric framework, we will briefly recall in Section 3 the geometry of consistent b.f.s. As the latter form a complex, we need to solve the approximation problem separately for each maximal simplicial component of such complex (Section 4). After gaining some insight from the analysis of the binary case (Section 5), we will proceed to solve the L_1 and L_2 consistent approximation problems in the general case in Section 6. We will finally comment and interpret our results.

2 Semantics of consistent belief functions

2.1 Consistent belief functions

We first recall the basic notions of the theory of evidence, and the definition of consistent belief functions in particular, to later discuss their semantics [19].

Definition 1 *A* basic probability assignment *(b.p.a.) on a finite set* (frame of discernment [19]) Θ *is a set function* $m_b : 2^\Theta \rightarrow [0,1]$ *on* $2^\Theta \doteq \{A \subseteq \Theta\}$ *s.t.*

$$m_b(\emptyset) = 0, \quad \sum_{A\subseteq\Theta} m_b(A) = 1, \quad m_b(A) \geq 0 \ \forall A \subseteq \Theta.$$

Subsets of Θ associated with non-zero values of m_b are called *focal elements* (f.e.), and their intersection *core*:

$$\mathcal{C}_b \doteq \bigcap_{A\subseteq\Theta: m_b(A)\neq 0} A.$$

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

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

A dual mathematical representation of the evidence encoded by a belief function b is the *plausibility function* (pl.f.) $pl_b : 2^\Theta \rightarrow [0,1]$, $A \mapsto pl_b(A)$ where

$$pl_b(A) \doteq 1 - b(A^c) = 1 - \sum_{B\subseteq A^c} m_b(B)$$

expresses the amount of evidence *not against* A.

In the theory of evidence a probability function is simply a special belief function assigning non-zero masses to singletons only (*Bayesian* b.f.): $m_b(A) = 0 \ |A| > 1$. *Consonant* belief functions are b.f.s whose f.e.s $A_1 \subset \cdots \subset A_m$ are nested. Consonant b.f.s always have a non-empty core, namely their smallest f.e. A_1. However, not all b.f.s whose core is non-empty are consonant.

Definition 3 *A belief function is said to be* consistent *if its core is non-empty.*

2.2 Semantics of consistent belief functions

Consistent belief functions (cs.b.f.s) form a significant class of b.f.s, for several reasons. On one side, they correspond to possibility distributions, and form therefore with consonant b.f.s the link between evidence and possibility theory. More importantly, though, they are the analogues of consistent, non-contradictory sets of propositions ("knowledge bases") in logics. As maintaining coherence along an inference process is highly desirable, the utility of an operator which maps arbitrary belief functions to consistent ones emerges. This is all the more valuable as several evidence combination rules, like Dempster's sum [8] and disjunctive combination [21] do not preserve consistency. To guarantee the consistency of the knowledge state a scheme like the following (where we use $\oplus$ to denote a valid combination rule) can be brought forward

$$\begin{array}{ccccc} b_1, b_2 & \rightarrow & b_1 \oplus b_2 & & \\ & & \downarrow & & \\ & & cs[b_1 \oplus b_2], \quad b_3 & \rightarrow & cs[b_1 \oplus b_2] \oplus b_3 \\ & & & & \downarrow \\ & & & & cs[cs[b_1 \oplus b_2] \oplus b_3] \end{array} \quad (1)$$

in which when new evidence is combined to yield a new belief state, the consistent transformation $cs[\cdot]$ is applied to ensure coherence.

2.3 Consistent b.f.s and possibility distributions

In possibility theory [9, 14], subjective probability is mathematically described by *possibility measures*, i.e. functions $Pos : 2^\Theta \to [0,1]$ such that $Pos(\emptyset) = 0$, $Pos(\Theta) = 1$ and $Pos\big(\bigcup_i A_i\big) = \sup_i Pos(A_i)$, for any family of subsets $\{A_i | A_i \in 2^\Theta, i \in I\}$, where I is an arbitrary set index.
Each measure Pos is uniquely characterized by a *possibility distribution* $\pi : \Theta \to [0,1]$, $\pi(x) \doteq Pos(\{x\})$, via the formula $Pos(A) = \sup_{x \in A} \pi(x)$.
A central role in the connection between possibility and evidence theory [20, 18, 14, 12, 23, 3] is played by consonant and consistent belief functions. On one side,

Proposition 1 *The plausibility function pl_b associated with a b.f. b is a possibility measure iff b is consonant.*

On the other, after calling *plausibility assignment* $\bar{pl}_b$ the restriction of the plausibility function to singletons $\bar{pl}_b(x) = pl_b(\{x\})$ it can be proven that [13, 5]

Proposition 2 *The plausibility assignment $\bar{pl}_b$ associated with a belief function b is the admissible possibility distribution of a possibility measure iff the b.f. b is consistent.*

Consistent b.f.s are then the counterparts of possibility distributions in the theory of evidence.
A different, powerful semantics comes in terms of consistent knowledge bases.

2.4 Consistent b.f.s as collections of coherent pieces of evidence

Belief functions are complex objects, in which sometimes contradictory bodies of evidence may coexist, as they may result from the fusion of possibly conflicting expert opinions and/or imprecise/corrupted measurements. In formal logics, the application of inference rules to inconsistent sets of assumptions or "knowledge bases" may lead to incompatible conclusions, depending on the subset of assumptions we start from.
A variety of approaches to solve this problem have been proposed. These include fragmenting the knowledge base into maximally consistent subsets, limiting the power of the formalism, or adopting non-classical semantics [17, 2]. Paris, on his side, tackles the problem by not assuming each proposition in the knowledge base as a fact, but by attributing to it a certain degree of belief [16]. This leads to something similar to a belief function.
A mechanism able to obtain a consistent knowledge base from an inconsistent one is therefore desirable. In the theory of evidence such a mechanism can be described as an operator

$$cs : \mathcal{B} \to \mathcal{CS}, \quad b \mapsto cs[b]$$

where $\mathcal{B}, \mathcal{CS}$ denote respectively the set of all b.f.s, and that of all cs.b.f.s.

2.5 Consistent belief functions and combination rules

Such a transformation acquires even more importance when we notice that most operators used to update/elicit evidence in the theory of evidence *do not preserve* consistency.

Definition 4 *The* orthogonal sum *or* Dempster's sum *of two belief functions b_1, b_2 is a new belief function $b_1 \oplus b_2$ with b.p.a.*

$$m_{b_1 \oplus b_2}(A) = \frac{\sum_{B \cap C = A} m_{b_1}(B)\, m_{b_2}(C)}{\sum_{B \cap C \neq \emptyset} m_{b_1}(B)\, m_{b_2}(C)},$$

where m_{b_i} denotes the b.p.a. associated with b_i.
Their disjunctive *combination is a new belief function $b_1 \cap b_2$ with b.p.a.*

$$m_{b_1 \cap b_2}(A) = \sum_{B \cap C = A} m_{b_1}(B) m_{b_2}(C).$$

Their conjunctive *combination is instead the b.f. $b_1 \cup b_2$ with b.p.a.*

$$m_{b_1 \cup b_2}(A) = \sum_{B \cup C = A} m_{b_1}(B) m_{b_2}(C).$$

Now, it is not difficult to prove that:

Proposition 3 *If b_1, b_2 are consistent then $b_1 \cup b_2$ is also consistent. On the other hand, if b_1, b_2 are consistent and their cores $\mathcal{C}_{b_1}$, $\mathcal{C}_{b_2}$ have non-empty intersection, then both $b_1 \oplus b_2$ and $b_1 \cap b_2$ are consistent with core $\mathcal{C}_{b_1 \cap b_2} = \mathcal{C}_{b_1} \cap \mathcal{C}_{b_2}$. Finally, if $\mathcal{C}_{b_1} \cap \mathcal{C}_{b_2} = \emptyset$ then $b_1 \oplus b_2$, $b_1 \cap b_2$ are* not *consistent.*

In other words, consistency is preserved by the conjunctive rule, the price to pay being increasing uncertainty as new evidence is combined, since the core of the belief state tends to Θ (complete ignorance). On the other side, both Dempster's rule and disjunctive combination preserve consistency only when the collection of focal elements of b_1 *and* b_2 is *already* consistent (i.e. any intersection $A \cap B$ of a f.e. A of b_1 and a f.e. B of b_2 is non-empty). As long as the new evidence is consistent with the existing one uncertainty is reduced. The price to pay is the loss of consistency in most cases.
The use of a consistent transformation in a reasoning process (1) would then guarantee consistency, while allowing the degree of uncertainty affecting our knowledge of the problem to decrease with time.

2.6 Making a belief function consistent

Consistent transformations can be built by solving a minimization problem of the form

$$cs[b] = \arg\min_{cs \in \mathcal{CS}} dist(b, cs) \quad (2)$$

where *dist* is some distance measure between belief functions, and $\mathcal{CS}$ denotes again the collection of all consistent b.f.s.
We call (2) the *consistent approximation problem.*
Plugging in different distance functions in (2) we get different consistent transformations.

In this paper we study what happens when using classical L_p norms in the approximation problem. As possibility measures are inherently related to the L_∞ norm (see above) cs.b.f.s live in a space linked to such a norm (Section 3). This leads to suppose that L_p-based approximations may indeed generate meaningful consistent transformations.

3 The simplicial complex of consistent belief functions

To solve the consistent approximation problem (2) we need to understand the structure of the space in which consistent belief functions live. We can then move forward and find the projection of b onto this space by minimizing the chosen distance.

3.1 The consistent complex

A belief function is determined by its $N-2$, $N = 2^{|\Theta|}$ belief values $\{b(A)\ \emptyset \subsetneq A \subsetneq \Theta\}$ (since $b(\emptyset) = 0$, $b(\Theta) = 1$ for all b.f.s). It can then be thought of as a vector of $\mathbb{R}^{N-2}$. The collection $\mathcal{B}$ of points of $\mathbb{R}^{N-2}$ which are b.f.s is a "simplex" (in rough words a higher-dimensional triangle), which we call *belief space*. $\mathcal{B}$ is the convex closure[1]

$$\mathcal{B} = Cl(b_A, \emptyset \subsetneq A \subseteq \Theta)$$

of the ("categorical") belief functions b_A assigning all the mass to a single event A: $m_b(A) = 1$, $m_b(B) = 0$ $\forall B \neq A$. In the belief space the vector $b \in \mathcal{B}$ which represents a belief function is the convex combination

$$b = \sum_{\emptyset \subsetneq A \subseteq \Theta} m_b(A) b_A \quad (3)$$

of the vectors b_A representing all the categorical belief functions.

[1]Here Cl denotes the convex closure operator: $Cl(b_1, ..., b_k) = \{b \in \mathcal{B} : b = \alpha_1 b_1 + \cdots + \alpha_k b_k, \sum_i \alpha_i = 1,\ \alpha_i \geq 0\ \forall i\}$.

The geometry of consistent belief functions can be described as a structure collection of simplices or *simplicial complex* [7]. More precisely, $\mathcal{CS}$ is the union

$$\mathcal{CS} = \bigcup_{x \in \Theta} Cl(b_A, A \ni x)$$

of the maximal simplices $Cl(b_A, A \ni x)$ formed by all the b.f.s with core containing a given element x of Θ.

3.2 Example: the binary case

As an example let us consider a frame of discernment formed by just two elements, $\Theta_2 = \{x, y\}$. In this very simple case each belief function $b : 2^{\Theta_2} \to [0,1]$ is completely determined by its belief values $b(x)$, $b(y)$ as $b(\Theta) = 1, b(\emptyset) = 0\ \forall b \in \mathcal{B}$.
We can then represent each b.f. b as the vector

$$[b(x) = m_b(x), b(y) = m_b(y)]'$$

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

$$m_b(x) \geq 0,\ m_b(y) \geq 0,\ m_b(x) + m_b(y) \leq 1$$

the set $\mathcal{B}_2$ of all the possible belief functions on Θ_2 is the triangle of Figure 1, whose vertices are the points $b_\Theta = [0,0]'$, $b_x = [1,0]'$, $b_y = [0,1]'$ which correspond 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 re-

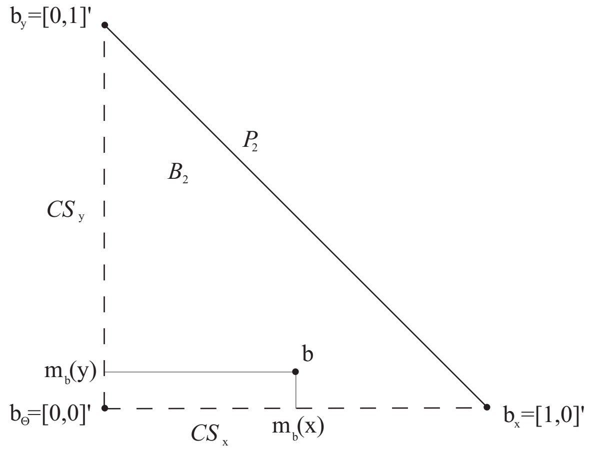

Figure 1: The belief space $\mathcal{B}$ for a binary frame is a triangle of $\mathbb{R}^2$ whose vertices are the categorical b.f.s focused on $\{x\}, \{y\}$ and Θ. The probability region is the segment $Cl(b_x, b_y)$, while all consistent b.f.s live in the union of the two segments $\mathcal{CS}_x = Cl(b_\Theta, b_x)$ and $\mathcal{CS}_y = Cl(b_\Theta, b_y)$.

gion $\mathcal{P}_2$ of all the Bayesian b.f.s on Θ_2 is the segment $Cl(b_x, b_y)$. In the binary case consistent belief functions can have as list of focal elements either $\{\{x\}, \Theta_2\}$

or $\{\{y\}, \Theta_2\}$. Therefore the space of cs.b.f.s $\mathcal{CS}_2$ is the union of two one-dimensional simplices (line segments):

$$\mathcal{CS}_2 = \mathcal{CS}_x \cup \mathcal{CS}_y = Cl(b_\Theta, b_x) \cup Cl(b_\Theta, b_y).$$

4 The L_p consistent approximation problem

4.1 Using norms of the L_p family

The geometry of the binary case hints to a strict relation between consistent belief functions and L_p norms. As the plausibility of all the elements of their core is

$$pl_b(x) = \sum_{A \supseteq \{x\}} m_b(A) = 1 \;\; \forall x \in \mathcal{C}_b,$$

the region of consistent b.f.s

$$\mathcal{CS} = \Big\{ b : \max_{x \in \Theta} pl_b(x) = 1 \Big\} = \Big\{ b : \|\bar{pl}_b\|_{L_\infty} = 1 \Big\}$$

is the set of b.f.s for which the L_∞ norm of the plausibility distribution is equal to 1. This reinforces the observation that cs.b.f.s correspond to possibility distributions (Section 2), which are in turn inherently related to L_∞.
It makes then sense to conjecture that the consistent transformation we obtain by picking as distance function in the approximation problem (2) one of the classical L_p norms

$$\begin{array}{rcl} \|b - b'\|_{L_1} & = & \sum_{A \subseteq \Theta} |b(A) - b'(A)|, \\ \|b - b'\|_{L_2} & = & \sqrt{\sum_{A \subseteq \Theta} (b(A) - b'(A))^2}, \\ \|b - b'\|_{L_\infty} & = & \max_{A \subseteq \Theta} \{|b(A) - b'(A)|\} \end{array}$$

will be meaningful.
When looking for a probabilistic approximation $p[b] = \arg\min_{p \in \mathcal{P}} dist(b, p)$ the use of L_p norms leads indeed to quite interesting results. The L_2 approximation produces the so-called "orthogonal projection" of b onto $\mathcal{P}$ [6], while, at least in the binary case, the set of L_1/L_∞ probabilistic approximations of b coincide with the set of probabilities dominating b:

$$\mathcal{P}[b] \doteq \{p \in \mathcal{P} : p(A) \geq b(A) \; \forall A \subseteq \Theta\}.$$

4.2 Approximation on a complex

As the consistent complex $\mathcal{CS}$ is a *collection* of linear spaces (better, simplices which generate a linear space) solving the problem (2) involves finding a number of partial solutions

$$cs^x_{L_p}[b] = \arg \min_{cs \in \mathcal{CS}_x} \|b - cs\|_{L_p} \qquad (4)$$

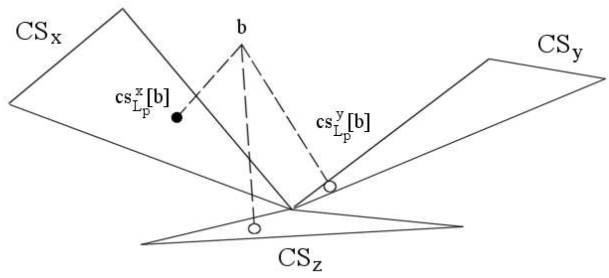

Figure 2: To minimize the distance of a point from a simplicial complex, we need to find all partial solutions (4) for all maximal simplices in the complex (empty circles), and later compare these partial solutions to select the global optimum (black circle).

(see Figure 2). Then, the distance of b from all such partial solutions has to be assessed in order to select a global optimal approximation.
In the rest of the paper we will apply this scheme to both the approximation problems associated with L_1 and L_2, respectively.

5 Approximation in the binary case

To get some insight on how to proceed in the general case, we will first consider the case study of a binary frame (Figure 3), and discuss how to approximate a belief function $b \in \mathcal{B}_2$ with a Bayesian or a consistent b.f. using an L_p norm. We will denote by

$$p_{L_p}[b] \doteq \arg \min_{p \in \mathcal{P}} \|b - p\|_{L_p}$$

the probability which minimizes the L_p distance from b. Analogously, we will use the notation

$$cs_{L_p}[b] \doteq \arg \min_{cs \in \mathcal{CS}} \|b - cs\|_{L_p}$$

for L_p consistent approximations.
In the Bayesian case we get

$$p_{L_2}[b] = \Big[m_b(x) + \frac{m_b(\Theta)}{2}, m_b(y) + \frac{m_b(\Theta)}{2} \Big]';$$

this probability is called *orthogonal projection* $\pi[b]$ of b onto $\mathcal{P}$ [6], and coincides with the pignistic function $BetP[b]$ [22, 4] in the binary case.
The L_1 solution $p_{L_1}[b]$, instead, is the whole set of probabilities "dominating" b [15], i.e.,

$$p_{L_1}[b] = \mathcal{P}[b] \doteq \big\{ p \in \mathcal{P} : p(A) \geq b(A) \; \forall A \subseteq \Theta \big\}. \quad (5)$$

Figure 3 illustrates the geometry of all L_p Bayesian and consistent approximations of a belief function b in the binary frame. We can notice that:

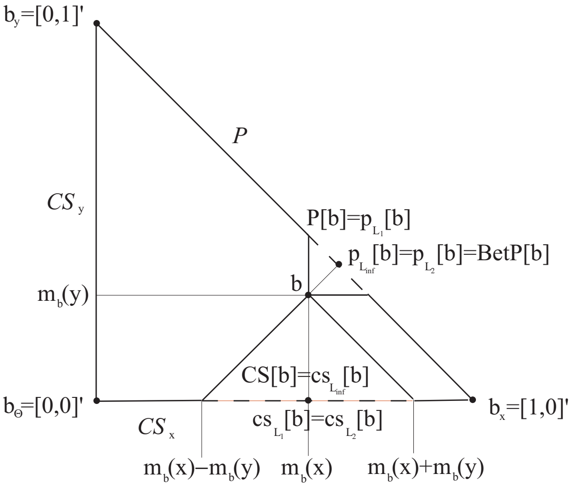

Figure 3: The dual behavior of Bayesian $p_{L_i}[b]$ and consistent $c_{L_i}[b]$ approximations of a b.f. b associated with the norms L_1, L_2, L_∞ is shown in the binary case.

1. the solution of the L_∞ approximation problem determines an entire set $\mathcal{CS}[b]$ of consistent b.f.s;
2. on the other hand, L_1/L_2 approximations on the same component $\mathcal{CS}_x$ of $\mathcal{CS}$ are point-wise and coincide;
3. the corresponding consistent transformation $cs^x_{L_2}[b]$ maps the original belief function b to a new b.f. with a focal element $A \cup \{x\}$ whenever A is a f.e. of b. The resulting b.p.a. is

$$\begin{aligned} m_{cs^x_{L_2}[b]}(x) &= \sum_{A:A\cup\{x\}=\{x\}} m_b(A) = m_b(x), \\ m_{cs^x_{L_2}[b]}(\Theta) &= \sum_{A:A\cup\{x\}=\Theta} m_b(A) \\ &= m_b(y) + m_b(\Theta). \end{aligned}$$

4. finally, the *global* L_1/L_2 consistent transformations also coincide, as they belong to the same component of the consistent complex ($\mathcal{CS}_x$ in the figure).

These facts (except the last point, which turns out to be an artifact of binary frames) are valid in the general case. Here we are going to focus on L_1/L_2 approximations.

6 Consistent L_1/L_2 approximations

6.1 Reducing the approximation problem to a linear system

In the case of an arbitrary frame a cs.b.f. $cs \in \mathcal{CS}_x$ is a solution of the L_2 *partial* approximation problem if $b - cs$ is orthogonal to all the generators $b_B - b_\Theta$ of the simplex $\mathcal{CS}_x = Cl(b_B, B \supseteq \{x\})$:

$$\langle b - cs, b_B - b_\Theta \rangle = \langle b - cs, b_B \rangle = 0 \quad \forall B \supseteq \{x\}$$

(as $b_\Theta = \mathbf{0}$ is the origin of $\mathbb{R}^{N-2}$, see binary example). We denote by $\alpha(A) \doteq m_{cs}(A)$ the b.p.a. of cs so that we can write each consistent belief function whose core contains $\{x\}$ as

$$cs = \sum_{A\supseteq\{x\}} \alpha(A) b_A$$

(by Equation (3)). After introducing the notation

$$\beta(A) \doteq m_b(A) - \alpha(A)$$

we can write $b - cs = \sum_{A\subsetneq\Theta} \beta(A) b_A$ and the orthogonality condition reads as

$$\Big\langle \sum_{A\subsetneq\Theta} \beta(A) b_A, b_B \Big\rangle = 0 \quad \forall B \supseteq \{x\}$$

i.e. (still for $\forall B \supseteq \{x\}$),

$$\Big\{ \sum_{A\supseteq\{x\}} \beta(A)\langle b_A, b_B\rangle + \sum_{A\not\supseteq\{x\}} m_b(A)\langle b_A, b_B\rangle = 0. \qquad (6)$$

The L_1 minimization problem reads instead as

$$\arg\min_{\vec{\alpha}} \Big\{ \sum_{A\supseteq\{x\}} \Big| \sum_{B\subseteq A} m_b(B) - \sum_{B\subseteq A, B\supseteq\{x\}} \alpha(B) \Big| \Big\} = \arg\min_{\vec{\beta}} \Big\{ \sum_{A\supseteq\{x\}} \Big| \sum_{B\subseteq A, B\supseteq\{x\}} \beta(B) + \sum_{B\subseteq A, B\not\supseteq\{x\}} m_b(B) \Big| \Big\}$$

which is clearly solved by setting all addenda to zero, obtaining the linear system:

$$\Big\{ \sum_{B\subseteq A, B\supseteq\{x\}} \beta(B) + \sum_{B\subseteq A, B\not\supseteq\{x\}} m_b(B) = 0 \; \forall A \supseteq \{x\}. \qquad (7)$$

6.2 Linear transformation

We are going to show here that the two minimization problems associated with the linear systems (6) and (7) coincide. The solution is indeed conserved due to the fact that the second linear system is obtained from the first one through a linear transformation.

Lemma 1 $\sum_{B\supseteq A} \langle b_B, b_C\rangle (-1)^{|B\setminus A|} = 1$ *if* $C \subseteq A$, *0 otherwise.*

Corollary 1 *The linear system (6) can be reduced to the system (7) through a linear transformation of rows:*

$$row_A \mapsto \sum_{B\supseteq A} row_B (-1)^{|B\setminus A|}. \qquad (8)$$

Proof. If we apply the linear transformation (8) to the system (6) we get

$$\sum_{B\supseteq A}\Big[\sum_{C\supseteq\{x\}}\beta(C)\langle b_B,b_C\rangle+\sum_{C\not\supset\{x\}}m_b(C)\langle b_B,b_C\rangle\Big]\cdot$$
$$\cdot(-1)^{|B\setminus A|}=\sum_{C\supseteq\{x\}}\beta(C)\sum_{B\supseteq A}\langle b_B,b_C\rangle(-1)^{|B\setminus A|}+$$
$$+\sum_{C\not\supset\{x\}}m_b(C)\sum_{B\supseteq A}\langle b_B,b_C\rangle(-1)^{|B\setminus A|}\quad\forall A\supseteq\{x\}.$$

Therefore by Lemma 1 we get

$$\sum_{C\supseteq\{x\},C\subseteq A}\beta_C+\sum_{C\not\supset\{x\},C\subseteq A}m_b(C)=0\qquad\forall A\supseteq\{x\}$$

i.e. the system of equations (7). □

6.3 Form of the solution

To obtain both the L_2 and the L_1 consistent approximations of b it then suffices to solve the system (7) associated with the L_1 norm.

Theorem 1 *The unique solution of the linear system (7) is given by*

$$\beta(A)=-m_b(A\setminus\{x\}).$$

Proof. We can prove it by substitution. System (7) becomes

$$-\sum_{B\subseteq A,B\supseteq\{x\}}m_b(B\setminus\{x\})+\sum_{B\subseteq A,B\not\supset\{x\}}m_b(B)=$$
$$=-\sum_{C\subseteq A\setminus\{x\}}m_b(C)+\sum_{B\subseteq A,B\not\supset\{x\}}m_b(B)=$$
$$=-\sum_{C\subseteq A\setminus\{x\}}m_b(C)+\sum_{C\subseteq A\setminus\{x\}}m_b(C)=0.\qquad\square$$

Therefore, according to what discussed in Section 4, the partial L_1/L_2 consistent approximations of b on the maximal component $\mathcal{CS}_x$ of the consistent complex have b.p.a.

$$\begin{aligned}m_{cs^x_{L_1}}(A)&=m_{cs^x_{L_2}}(A)=\alpha(A)=m_b(A)-\beta(A)\\&=m_b(A)+m_b(A\setminus\{x\})\end{aligned}$$

for all events A such that $\{x\}\subseteq A\subsetneq\Theta$.
The value of $\alpha(\Theta)$ can be obtained by normalization:

$$\begin{aligned}\alpha(\Theta)&=1-\sum_{\{x\}\subseteq A\subsetneq\Theta}\alpha(A)\\&=1-\sum_{\{x\}\subseteq A\subsetneq\Theta}m_b(A)+m_b(A\setminus\{x\})\\&=1-\sum_{\{x\}\subseteq A\subsetneq\Theta}m_b(A)-\sum_{\{x\}\subseteq A\subsetneq\Theta}m_b(A\setminus\{x\})\\&=1-\sum_{A\neq\Theta,\{x\}^c}m_b(A)=m_b(\{x\}^c)+m_b(\Theta)\end{aligned}$$

as $B\not\supset\{x\}$ iff $B=A\setminus\{x\}$ for $A=B\cup\{x\}$.

Corollary 2 *The partial L_1 and L_2 consistent approximations of a belief function b with b.p.a. m_b onto the component $\mathcal{CS}_x$ of the consistent complex coincide. They have b.p.a.*

$$m_{cs^x_{L_1}}(A)=m_{cs^x_{L_2}}(A)=m_b(A)+m_b(A\setminus\{x\})$$

$\forall x\in\Theta$, and for all A s.t. $\{x\}\subseteq A\subseteq\Theta$.

6.4 Partial solutions as focused consistent transformations

The basic probability assignment of the L_1/L_2 consistent approximations of b has an elegant expression. It also has a straightforward interpretation: to get a consistent b.f. focused on a singleton x, the mass contribution of all the events B such that $B\cup\{x\}=A$ coincide is assigned indeed to A. But there are just two such events: A itself, and $A\setminus\{x\}$.
As an example, the partial consistent approximation of a belief function on a frame $\Theta=\{x,y,z,w\}$ with core $\{x\}$ is illustrated in Figure 4. The b.f. with focal

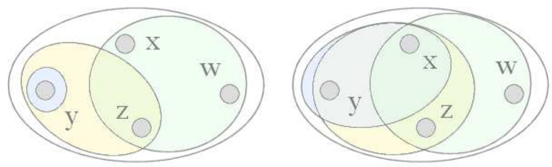

Figure 4: A belief function (left) and its L_1/L_2 consistent approximation with core $\{x\}$ (right).

elements $\{y\}$, $\{y,z\}$, and $\{x,z,w\}$ is transformed by the map

$$\begin{array}{lll}\{y\}&\mapsto&\{x\}\cup\{y\}=\{x,y\},\\\{y,z\}&\mapsto&\{x\}\cup\{y,z\}=\{x,y,z\},\\\{x,z,w\}&\mapsto&\{x\}\cup\{x,z,w\}=\{x,z,w\}\end{array}$$

into the consistent b.f. with focal elements $\{x,y\}$, $\{x,y,z\}$, and $\{x,z,w\}$ and the same b.p.a.

Partial solutions to the L_1/L_2 consistent approximation problem turn out to be related to classical *inner consonant approximations* of a belief function b, i.e. the set of consonant b.f.s such that $c(A)\geq b(A)$ $\forall A\subseteq\Theta$ (or equivalently $pl_c(A)\leq pl_b(A)$ $\forall A$).
Dubois and Prade [10] proved indeed that such an approximation exists iff b is consistent. However, when b is *not* consistent a "focused consistent transformation" can be applied to get a new belief function b' such that

$$m'(A\cup x_i)=m(A)\qquad\forall A\subseteq\Theta$$

and x_i is the element of Θ with highest plausibility.
Theorem 1 and Corollary 2 state that the L_1/L_2 consistent approximation onto each component $\mathcal{CS}_x$ of $\mathcal{CS}$ generates the consistent transformation focused on x.

6.5 Global optimal solution for L_1

To find the *global* consistent approximation of b we need to work out which of the partial approximations $cs^x_{L_{1/2}}[b]$ has minimal distance from b. To do so we need to find

$$\arg\min_x \|b - cs^x_{L_{1/2}}[b]\|.$$

The L_1 distance of b from $\mathcal{CS}_x$ can be computed as

$$\begin{array}{l} \|b - cs^x_{L_1}[b]\|_{L_1} = \displaystyle\sum_{A\subseteq\Theta} |b(A) - cs^x_{L_1}[b](A)| \\ = \displaystyle\sum_{A\not\supseteq\{x\}} |b(A) - 0| + \sum_{A\supseteq\{x\}} \Big|b(A) - \sum_{B\subseteq A, B\supseteq\{x\}} \alpha(B)\Big| \\ = \displaystyle\sum_{A\not\supseteq\{x\}} b(A) + \sum_{A\supseteq\{x\}} \Big|\sum_{B\subseteq A} m_b(B) + \\ - \displaystyle\sum_{B\subseteq A, B\supseteq\{x\}} \big(m_b(B) + m_b(B\setminus\{x\})\big)\Big| \\ = \displaystyle\sum_{A\not\supseteq\{x\}} b(A) + \sum_{A\supseteq\{x\}} \Big|\sum_{B\subseteq A, B\not\supseteq\{x\}} m_b(B) + \\ - \displaystyle\sum_{B\subseteq A, B\supseteq\{x\}} m_b(B\setminus\{x\})\Big| = \sum_{A\not\supseteq\{x\}} b(A) + \\ + \displaystyle\sum_{A\supseteq\{x\}} \Big|\sum_{C\subseteq A\setminus\{x\}} m_b(C) - \sum_{C\subseteq A\setminus\{x\}} m_b(C)\Big| \\ = \displaystyle\sum_{A\not\supseteq\{x\}} b(A) = \sum_{A\subseteq\{x\}^c} b(A). \end{array} \quad (9)$$

Immediately,

Theorem 2 *The global optimal L_1 consistent approximation of any belief function b is given by*

$$cs_{L_1}[b] \doteq \arg\min_{cs\in\mathcal{CS}} \|b - cs^x_{L_1}[b]\| = cs^{\hat{x}}_{L_1}[b]$$

i.e. the partial approximation associated with the element $\hat{x}$ which minimizes (9):

$$\hat{x} = \arg\min_x \Big\{ \sum_{A\subseteq\{x\}^c} b(A), x\in\Theta \Big\}.$$

6.6 A counterexample

In the binary case (Figure 3) the condition of Theorem 2 reduces to

$$\begin{array}{rcl} \hat{x} &=& \arg\min_x \displaystyle\sum_{A\subseteq\{x\}^c} b(A) = \arg\min_x m_b(\{x\}^c) \\ &=& \arg\max_x pl_b(x) \end{array}$$

and the global approximation falls on the component of the consistent complex associated with the element of *maximal plausibility*.
Unfortunately, this is not generally the case for arbitrary frames of discernment Θ. Let us see this in a simple counterexample. Let us first write

$$\begin{array}{l} \displaystyle\sum_{A\subseteq\{x\}^c} b(A) = \sum_{A\subseteq\{x\}^c}\sum_{B\subseteq A} m_b(B) = \sum_{B\subseteq\{x\}^c} m_b(B)\cdot \\ \cdot|\{A\subseteq\{x\}^c : A\supseteq B\}| = \displaystyle\sum_{B\subseteq\{x\}^c} m_b(B)\cdot 2^{|\{x\}^c|-|B|}. \end{array} \quad (10)$$

Now, consider a belief function on a frame $\Theta = \{x_1, ..., x_n\}$ of cardinality n, with just two focal elements:

$$\begin{array}{rcl} m_b(x_1) &=& m_x, \\ m_b(\{x_1\}^c) &=& m_b(\{x_2, ..., x_n\}) = 1 - m_x. \end{array}$$

If $m_x < 1/2$ all $y \neq x_1$ have maximal plausibility, as $pl_b(x_1) = 1 - b(\{x_1\}^c) = m_x$, while $pl_b(y) = 1 - m_x$ for all $y \neq x$. However, according to (10),

$$\begin{array}{rcl} \|b - cs^{x_1}_{L_1}[b]\|_{L_1} &=& \displaystyle\sum_{A\subseteq\{x_1\}^c} b(A) \\ &=& (1-m_x)2^{n-1-(n-1)} = 1 - m_x, \end{array}$$

where $n = |\Theta|$, while

$$\begin{array}{rcl} \|b - cs^{y}_{L_1}[b]\|_{L_1} &=& \displaystyle\sum_{A\subseteq\{y\}^c} b(A) \\ &=& m_x 2^{n-1-1} = m_x 2^{n-2} \end{array}$$

$\forall y \neq x$. But when

$$m_x 2^{n-2} \geq 1 - m_x \equiv n \geq 2 + \log_2\Big(\frac{1-m_x}{m_x}\Big)$$

we have that

$$\|b - cs^{x_1}_{L_1}[b]\|_{L_1} \leq \|b - cs^{y}_{L_1}[b]\|_{L_1} \quad \forall y \neq x_1,$$

and therefore *the global L_1 consistent approximation can fall on a component not associated with the maximal plausibility element.*

6.7 Global optimal solution for L_2

In the L_2 case we get

$$\begin{array}{l} \|b - cs^x_{L_2}[b]\|^2 = \displaystyle\sum_{A\subseteq\Theta} \Big(b(A) - cs^x_{L_2}[b](A)\Big)^2 = \\ = \displaystyle\sum_{A\subseteq\Theta} \Big[\sum_{B\subseteq A} m_b(B) - \sum_{B\subseteq A, B\supseteq\{x\}} \alpha(B)\Big]^2 = \\ = \displaystyle\sum_{A\subseteq\Theta} \Big[\sum_{B\subseteq A} m_b(B) - \sum_{B\subseteq A, B\supseteq\{x\}} m_b(B) + \\ - \displaystyle\sum_{B\subseteq A, B\supseteq\{x\}} m_b(B\setminus\{x\})\Big]^2 = \\ = \displaystyle\sum_{A\not\supseteq\{x\}} \big(b(A)\big)^2 + \sum_{A\supseteq\{x\}} \Big[\sum_{B\subseteq A, B\not\supseteq\{x\}} m_b(B) + \\ - \displaystyle\sum_{B\subseteq A, B\supseteq\{x\}} m_b(B\setminus\{x\})\Big]^2 = \sum_{A\not\supseteq\{x\}} \big(b(A)\big)^2 + \\ + \displaystyle\sum_{A\supseteq\{x\}} \Big[\sum_{C\subseteq A\setminus\{x\}} m_b(C) - \sum_{C\subseteq A\setminus\{x\}} m_b(C)\Big]^2 \end{array}$$

so that, in analogy with the L_1 case,

$$\|b - cs^x_{L_2}[b]\|^2 = \sum_{A \subseteq \{x\}^c} (b(A))^2.$$

Theorem 3 *The global optimal L_2 consistent approximation of any belief function b is given by*

$$cs_{L_2}[b] \doteq \arg\min_{cs \in \mathcal{CS}} \|b - cs^x_{L_2}[b]\| = cs^{\hat{x}}_{L_2}[b]$$

i.e. the partial approximation associated with the element

$$\hat{x} = \arg\min_x \Big\{ \sum_{A \subseteq \{x\}^c} (b(A))^2, x \in \Theta \Big\}.$$

Other simple counterexamples show that the global L_2 consistent approximation can fall on a component not associated with the maximal plausibility element.

7 Comments and conclusions

Belief functions represent coherent knowledge bases in the theory of evidence. As consistency is not preserved by most operators used to update or elicit evidence, the use of a consistent transformation in conjunction with those combinations rules can be desirable. Consistent transformations are strictly related to the problem of approximating a generic belief function with a consistent one.

In this paper we solved the instance of the consistent approximation problem we obtain when measuring distances between uncertainty measures by means of the classical L_p norms. This makes sense as cs.b.f.s live in a simplicial complex defined in terms of the L_∞ norms, and correspond to possibility distributions. A partial approximation for each component of the complex has to be found. The conclusions of this study are the following:
1. partial L_1/L_2 approximations coincide on each component of the consistent complex;
2. such partial approximation turns out to be the consistent transformation focused on the given element of the frame;
3. the corresponding global solutions have not in general as core the maximal plausibility element, and may lie in general on different components of $\mathcal{CS}$.

The interpretation of the polytope of all L_∞ solutions is worth to be fully investigated in the near future, in the light of the intuition provided by the binary case. In particular its clear analogy with the polytope of consistent probabilities will be interesting matter to study. A natural continuation of this line of research is obviously the solution of the L_p approximation problem for consonant belief functions, as counterparts of possibility measures in the theory of evidence. That will complete our understanding of the relation between geometric norms and evidence consistency.

Proof of Lemma 1

We first note that, by definition of dogmatic belief function b_A (Section 3),

$$\langle b_B, b_C \rangle = \sum_{D \supseteq B, C; D \neq \Theta} 1 = \sum_{E \subsetneq (B \cup C)^c} 1 = 2^{|(B \cup C)^c|} - 1.$$

Hence $\sum_{B \supseteq A} \langle b_B, b_C \rangle (-1)^{|B \setminus A|} =$

$$\begin{aligned} &= \sum_{B \supseteq A} (2^{|(B \cup C)^c|} - 1)(-1)^{|B \setminus A|} \\ &= \sum_{B \supseteq A} 2^{|(B \cup C)^c|}(-1)^{|B \setminus A|} - \sum_{B \supseteq A} (-1)^{|B \setminus A|} \\ &= \sum_{B \supseteq A} 2^{|(B \cup C)^c|}(-1)^{|B \setminus A|}, \end{aligned}$$

as

$$\sum_{B \supseteq A} (-1)^{|B \setminus A|} = \sum_{k=0}^{|B \setminus A|} 1^{|A^c| - k}(-1)^k = 0$$

for Newton's binomial:

$$\sum_{k=0}^{n} p^k q^{n-k} = (p+q)^n. \tag{11}$$

Now, as both $B \supseteq A$ and $C \supseteq A$ the set B can be

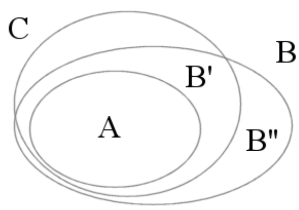

Figure 5: Decomposition of B into $A + B' + B''$ in the proof of Lemma 1.

decomposed into the disjoint sum

$$B = A + B' + B''$$

where

$$\emptyset \subseteq B' \subseteq C \setminus A, \quad \emptyset \subseteq B'' \subseteq (C \cup A)^c$$

(see Figure 5), so that the above quantity can be written as

$$\sum_{\emptyset \subseteq B' \subseteq C \setminus A} \sum_{\emptyset \subseteq B'' \subseteq (C \cup A)^c} 2^{|(A \cup C)|^c - |B''|}(-1)^{|B'| + |B''|} = \sum_{\emptyset \subseteq B' \subseteq C \setminus A} (-1)^{|B'|} \sum_{\emptyset \subseteq B'' \subseteq (C \cup A)^c} (-1)^{|B''|} 2^{|(A \cup C)|^c - |B''|}$$

where

$$\sum_{\emptyset\subseteq B''\subseteq(C\cup A)^c} (-1)^{|B''|}2^{|(A\cup C)|^c-|B''|} = [2+(-1)]^{|(A\cup C)|^c}$$

$= 1^{|(A\cup C)|^c} = 1$, again for Newton's binomial (11). The desired quantity becomes

$$\sum_{\emptyset\subseteq B'\subseteq C\setminus A} (-1)^{|B'|}$$

which is nil for $C \setminus A \neq \emptyset$, equal to 1 when $C \setminus A = \emptyset$, i.e. $C \subseteq A$.

References

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

[2] D. Batens, C. Mortensen, and G. Priest, *Frontiers of paraconsistent logic*, Studies in logic and computation (J.P. Van Bendegem, ed.), vol. 8, Research Studies Press, 2000.

[3] L. Caro and A. Babak Nadjar, *Generalization of the Dempster-Shafer theory: a fuzzy-valued measure*, IEEE Transactions on Fuzzy Systems **7** (1999), 255–270.

[4] B.R. Cobb and P.P. Shenoy, *A comparison of Bayesian and belief function reasoning*, Information Systems Frontiers **5** (2003), no. 4, 345–358.

[5] F. Cuzzolin, *An interpretation of consistent belief functions in terms of simplicial complexes*, submitted to Information Sciences (2007).

[6] F. Cuzzolin, *Two new Bayesian approximations of belief functions based on convex geometry*, IEEE Trans. on Systems, Man, and Cybernetics - Part B **37** (2007), no. 4, 993–1008.

[7] F. Cuzzolin, *An interpretation of consistent belief functions in terms of simplicial complexes*, Proc. of ISAIM'08, 2008.

[8] A.P. Dempster, *A generalization of Bayesian inference*, Journal of the Royal Stat. Soc., Series B **30** (1968), 205–247.

[9] D. Dubois and H. Prade, *Possibility theory*, Plenum Press, New York, 1988.

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

[11] B.A. Dubrovin, S.P. Novikov, and A.T. Fomenko, *Sovremennaja geometrija. metody i prilozenija*, Nauka, Moscow, 1986.

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

[13] C. Joslyn, *Towards an empirical semantics of possibility through maximum uncertainty*, Proc. IFSA 1991 (R. Lowen and M. Roubens, eds.), vol. A, 1991, pp. 86–89.

[14] G. J. Klir, W. Zhenyuan, and D. Harmanec, *Constructing fuzzy measures in expert systems*, Fuzzy Sets and Systems **92** (1997), 251–264.

[15] H. Kyburg, *Bayesian and non-Bayesian evidential updating*, Artificial Intelligence **31** (1987), no. 3, 271–294.

[16] J.B. Paris, D. Picado-Muino, and M. Rosefield, *Information from inconsistent knowledge: A probability logic approach*, Interval / Probabilistic Uncertainty and Non-classical Logics, Advances in Soft Computing, vol. 46, Springer-Verlag, Berlin - Heidelberg, 2008.

[17] G. Priest, R. Routley, and J. Norman, *Paraconsistent logic: Essays on the inconsistent*, Philosophia Verlag, 1989.

[18] C. Roemer and A. Kandel, *Applicability analysis of fuzzy inference by means of generalized Dempster-Shafer theory*, IEEE Transactions on Fuzzy Systems **3** (1995), no. 4, 448–453.

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

[20] Ph. Smets, *The transferable belief model and possibility theory*, NAFIPS-90 (Kodratoff Y., ed.), 1990, pp. 215–218.

[21] Ph. Smets, *Belief functions : the disjunctive rule of combination and the generalized Bayesian theorem*, International Journal of Approximate Reasoning **9** (1993), 1–35.

[22] Ph. Smets and R. Kennes, *The transferable belief model*, Artificial Intelligence **66** (1994), 191–234.

[23] R. R. Yager, *Class of fuzzy measures generated from a Dempster-Shafer belief structure*, International Journal of Intelligent Systems **14** (1999), 1239–1247.