# Approximating the Combination of Belief Functions using the Fast Möbius Transform in a Coarsened Frame

Thierry Denœux*[1] and Amel Ben Yaghlane†

* Université de Technologie de Compiègne
UMR CNRS 6599 Heudiasyc
Centre de Recherches de Royallieu
BP 20529 - F-60205 Compiègne cedex - France
email: Thierry.Denoeux@utc.fr
† Institut Supérieur de Gestion de Tunis,
41 Avenue de la liberté, cité Bouchoucha, 2000 Le Bardo - Tunis - Tunisia

March 26, 2002

[1]Corresponding author.

**Abstract**

A method is proposed for reducing the size of a frame of discernment, in such a way that the loss of information content in a set of belief functions is minimized. This method may be seen as a hierarchical clustering procedure applied to the columns of a binary data matrix, using a particular dissimilarity measure. It allows to compute approximations of the mass functions, which can be combined efficiently in the coarsened frame using the Fast Möbius Transform algorithm, yielding inner and outer approximations of the combined belief function.

**Keywords:** Belief Functions, Dempster-Shafer Theory, Evidence Theory, Approximation, Evidential Reasoning.

# 1 Introduction

The Dempster-Shafer theory of Belief Functions (BF's) is now widely accepted as a rich and flexible framework for representing and reasoning with imperfect information. The concept of belief function subsumes those of probability and possibility measures, making the theory very general. Situations of weak knowledge and heterogeneous information sources are easily modeled, making it quite suitable in many application domains such as medical diagnosis, sensor fusion and pattern recognition [18].

This generality, however, has a cost in terms of computational complexity. A BF (or, equivalently, a mass function) assigns a number to each of the $2^n$ subsets of the frame of discernment $\Omega$ ($|\Omega| = n$), with $2^n - 1$ degrees of freedom, which is much larger than what is needed to specify a probability or a possibility measure. Although BF's as elicited from experts or inferred from statistical data are usually constrained to be of a simple form, the fusion of several BF's using the Dempster's rule of combination almost inevitably increases the number of focal elements (i.e., subsets of $\Omega$ with a positive mass of belief), resulting in high storage and computational requirements for large-scale problems.

The algorithmic complexity of combining several BF's has been studied from a theoretical point of view by Orponen [11], who proved that the problem is $\#P$ complete. Recently, Wilson [20] provided a very complete review of algorithmic issues related to the manipulation of BF's. Currently, two algorithms exist for computing the conjunctive combination $m_1 \circledcirc m_2$ of two mass functions $m_1$ and $m_2$ (similar methods hold for the disjunctive combination):

- the mass-based algorithm, initially sketched by Shafer [15], involves considering each focal element $A$ of $m_1$, each focal element $B$ of $m_2$, and assigning the mass $m_1(A)m_2(B)$ to the set $A \cap B$. Using this method, the combination can be performed in time proportional to $n|\mathcal{F}(m_1)||\mathcal{F}(m_2)|$, where $\mathcal{F}(m_k)$ denotes the number of focal elements of $m_k$ ($k = 1, 2$). The time needed for the combination of $K$ BF's $m_1, \ldots, m_K$ depends on the particular structure of the mass functions, and can be in the worst case exponential in $\min(n, K)$, as shown by Wilson [20, page 442].

- the Fast Möbius Transform (FMT) method [9, 8] converts each mass function $m_k$ into its associated commonality function $q_k$; the product of these functions is computed, and the result is converted back into a mass function. The algorithm takes time proportional to $Kn^2 2^n$ (see the discussion in [20, page 444]).

The choice of one of these methods depends on the structure of the mass functions. As remarked by Wilson, if the number of focal elements of the combined belief function is much smaller than $2^n$, then the mass-based method is likely to be faster. However, this is generally not known in advance. If one of the BF's has a number of focal elements close to $2^n$, then the FMT method is likely to be faster. However, this method requires to store $2^n$ reals for representing each belief function, and it thus becomes impractical when $\Omega$ has more than 15 to 20 elements.

When the combination of several BF's cannot be computed exactly, one has to resort to stochastic or deterministic approximation procedures [20]. Monte-Carlo algorithms can be very efficient for computing values of combined belief for a small number of subsets of $\Omega$: for instance, the commonality-based importance sampling algorithm described in [20, page 458] is approximately linear in the size of the frame. If, however, one is interested in the whole belief function (as it is usually the case in real applications), then stochastic methods are no longer feasible for large $\Omega$, and deterministic approaches have to be used. Since the mass-based method for combining BF's is the most widely used, most deterministic methods (which are exclusively considered here) have been designed with the aim of reducing the number of focal elements. This is true, in particular, for the summarization method initially introduced by Lowrance et al. [6], and for the more sophisticated methods proposed subsequently [19] [1] [5] [14] [2].

In this paper, a different approach is investigated. Instead of reducing the number of focal elements, we propose to *reduce the size of the frame of discernment*, which can be expected to drastically decrease the computing time of the FMT combination method, and make it applicable to find reasonable approximations in the case of large-size problems. Given a set of BF's, we propose to find a coarsening of the frame $\Omega$ that will preserve as much as possible of the information content of the belief functions. This approach allows to compute inner and outer approximations, from which lower and upper bounds for the combined belief values can be derived.

The following section summarizes the background definitions and results needed in the sequel. Our approximation method is then described in Section 3, and simulation results are presented in Section 4.

# 2 Background

## 2.1 Basic Concepts

Only the main concepts of evidence theory are summarized here. More details can be found in, e.g., Refs. [15] and [17]. Let $\Omega$ denote a finite set called the frame of discernment. A mass function, or *basic belief assignment*(bba) is a function $m : 2^\Omega \to [0,1]$ verifying:

$$\sum_{A \subseteq \Omega} m(A) = 1. \tag{1}$$

Each mass of belief $m(A)$ measures the amount of belief that is exactly committed to $A$. A bba $m$ such that $m(\emptyset) = 0$ is said to be normal. This condition will not be imposed here. The subsets $A$ of $\Omega$ such that $m(A) > 0$ are called *focal elements* of $m$. Let $\mathcal{F}(m) \subseteq 2^\Omega$ denote the set of focal elements of $m$.

The *belief function* induced by $m$ is a function $\text{bel} : 2^\Omega \to [0,1]$, defined as:

$$\text{bel}(A) = \sum_{\emptyset \neq B \subseteq A} m(B) \tag{2}$$

for all $A \subseteq \Omega$. bel($A$) represents the amount of support given to $A$.

The *plausibility function* associated with a bba $m$ is a function $\text{pl} : 2^\Omega \to [0,1]$, defined as:

$$\text{pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \qquad \forall A \subseteq \Omega. \tag{3}$$

pl($A$) represents the potential amount of support that *could be* given to $A$.

Given two bba's $m_1$ and $m_2$ defined over the same frame of discernment $\Omega$ and induced by two distinct pieces of evidence, we can combine them in two ways using the conjunctive or the disjunctive rules of combination [17] defined, respectively, as:

$$(m_1 \textcircled{\cap} m_2)(A) = \sum_{B \cap C = A} m_1(B) m_2(C) \tag{4}$$

$$(m_1 \bigcirc\!\!\!\!\!\cup\, m_2)(A) = \sum_{B \cup C = A} m_1(B) m_2(C) \tag{5}$$

for all $A \subseteq \Omega$. The choice of one of these combination rules is related to the reliability of the two sources. If we know that both sources of information are fully reliable, then the corresponding bba's should be combined conjunctively. In contrast, if we only know that at least one of the two sources is reliable, then the disjunctive rule must be used.

The conjunctive and disjunctive rules can be conveniently expressed by means of the commonality function $q$ and the implicability function $b$, defined, respectively, as

$$q(A) = \sum_{A \subseteq B} m(B) \tag{6}$$

and

$$b(A) = \mathrm{bel}(A) + m(\emptyset) \tag{7}$$

for all $A \subseteq \Omega$. If $q_1 \bigcirc\!\!\!\!\!\cap\, q_2$ denotes the commonality function associated to $m_1 \bigcirc\!\!\!\!\!\cap\, m_2$, and $b_1 \bigcirc\!\!\!\!\!\cup\, b_2$ denotes the implicability function associated to $m \bigcirc\!\!\!\!\!\cup\, m_2$, we have the following simple relations [16]:

$$q_1 \bigcirc\!\!\!\!\!\cap\, q_2 = q_1 q_2 \tag{8}$$

$$b_1 \bigcirc\!\!\!\!\!\cup\, b_2 = b_1 b_2 \ . \tag{9}$$

The importance of this result arises from the fact that functions $m$, $q$ and $b$ (as well as bel and pl) are equivalent representations, in the sense that, given any of these functions, it is possible to recover all the others. The conversion between these functions can be efficiently done using the FMT algorithm [9, 8] in time proportional to $n^2 2^n$ [20]. Relations (8) and (9) provide the basis for the FMT-based method for combining bba's, which consists in transforming the bba's to $q$ or $b$, computing the product, and converting back the result into a mass function. In contrast, the more traditional mass-based approach relies exclusively on Eqs (4) and (5).

## 2.2 Inclusion of Belief Functions

Another notion of interest is that of strong inclusion of bba's [3]. Let $m$ and $m'$ be two BS's with focal elements $\mathcal{F}(m) = \{F_1, \ldots, F_p\}$ and $\mathcal{F}(m') = \{F'_1, \ldots, F'_{p'}\}$. Then, $m$ is said to be strongly included in $m'$, or to be a *specialization* of $m'$ (which is noted

$m \subseteq m'$), iff there exists a non-negative matrix $W$ with entries $w_{ij}$ ($i = 1, \ldots, p; j = 1, \ldots, p'$) such that

$$\sum_{j=1}^{p'} w_{ij} = m(F_i), \quad i = 1, \ldots, p, \tag{10}$$

$$\sum_{i=1}^{p} w_{ij} = m'(F'_j), \quad j = 1, \ldots, p' \tag{11}$$

and $w_{ij} > 0 \Rightarrow F_i \subseteq F'_j$. The relationship between $m$ and $m'$ may be seen as a transfer of mass from each focal element $F_i$ of $m$ to supersets $F'_j \supseteq F_i$, the quantity $w_{ij}$ denoting the part of $m(F_i)$ transferred to $F'_j$.

In the context of belief function approximation, a bba $\underline{m}$ (resp. $\overline{m}$) is called a strong inner (resp. outer) approximation of a bba $m$ if it satisfies $\underline{m} \subseteq m$ (resp. $m \subseteq \overline{m}$). Approximating a bba $m$ by a pair $(\underline{m}, \overline{m})$ of strong inner and outer approximations is very interesting, because it is allows to obtain lower and upper bounds for functions $pl$, $b$ and $q$ when combining several BF's, as shown by the following two propositions [3, 2].

**Proposition 1** *Let $(\underline{m}, \overline{m})$ a pair of strong inner and outer approximation of a bba $m$, and let $(\underline{pl}, \overline{pl})$, $(\underline{q}, \overline{q})$ and $(\underline{b}, \overline{b})$ the corresponding approximations of $pl$, $q$ and $b$, respectively. We have:*

$$\underline{pl} \leq pl \leq \overline{pl}$$

$$\underline{q} \leq q \leq \overline{q}$$

$$\overline{b} \leq b \leq \underline{b} \ .$$

**Remark 1** *In the case of subnormal bbas's, we do not have in general $\underline{bel} \leq bel \leq \overline{bel}$, because the mass given to the empty set is not counted in the calculation of belief values. However, a bracketing of bel may still be obtained [2] by noticing that*

$$bel(A) = pl(\Omega) - pl(\overline{A}) \quad \forall A \in [0,1]^{\Omega},$$

*where $\overline{A}$ denotes the complement of $A$. From this expression we can derive the following inequalities for all $A \in [0,1]^{\Omega}$:*

$$\max[0, \underline{pl}(\Omega) - \overline{pl}(\overline{A})] \leq bel(A) \leq \overline{pl}(\Omega) - \underline{pl}(\overline{A}) \ . \tag{12}$$

*Similarly, lower and upper bounds for normalized belief and plausibility values based on inner and outer approximations are given in [2].*

**Proposition 2** *Let $(\underline{m}_1, \overline{m}_1)$ and $(\underline{m}_2, \overline{m}_2)$ be strong approximations of two bba's $m_1$ and $m_2$, respectively. We have:*

$$\underline{m}_1 \bigcirc\!\!\!\!\!\cap\; \underline{m}_2 \subseteq m_1 \bigcirc\!\!\!\!\!\cap\; m_2 \subseteq \overline{m}_1 \bigcirc\!\!\!\!\!\cap\; \overline{m}_2$$

$$\underline{m}_1 \bigcirc\!\!\!\!\!\cup\; \underline{m}_2 \subseteq m_1 \bigcirc\!\!\!\!\!\cup\; m_2 \subseteq \overline{m}_1 \bigcirc\!\!\!\!\!\cup\; \overline{m}_2 \ .$$

Methods for constructing strong inner and outer approximations have been proposed by Dubois and Prade [4] in a possibilistic setting, and by Denœux [2] using an approach based on the clustering of focal elements.

### 2.3 Coarsenings and Refinements

#### Main definitions

In applying the BF framework to a real-world problem, the definition of the frame of discernment is a crucial step. As remarked by Shafer [15], the degree of "granularity" of the frame is always, to some extent, a matter of convention, as any element $\omega$ of $\Omega$ representing a "state of nature" can always be split into several possibilities. Hence, it is fundamental to examine how a BF defined on a frame may be expressed in a finer or, conversely, in a coarser frame.

Let $\Omega$ and $\Theta$ denote two finite sets. A mapping $\rho : 2^\Theta \rightarrow 2^\Omega$ is called a *refining* if it verifies the following properties:

1. The set $\{\rho(\{\theta\}), \theta \in \Theta\} \subseteq 2^\Omega$ is a partition of $\Omega$.

2. For all $A \subseteq \Theta$, we have

$$\rho(A) = \bigcup_{\theta \in A} \rho(\{\theta\}) \ . \tag{13}$$

Following the terminology introduced by Shafer, the set $\Theta$ is then called a *coarsening* of $\Omega$, and $\Omega$ is called a *refinement* of $\Theta$.

Note that defining a coarsening of a frame $\Omega$ is formally equivalent to defining a partition of $\Omega$. Let $\Theta$ be such a partition. The function $\rho : 2^\Theta \rightarrow 2^\Omega$ such that

$\rho(\{\theta\}) = \theta$ for all $\theta \in \Theta$, and verifying (13) is a refining of $\Theta$, and $\Theta$ is a coarsening of $\Omega$.

**Example 1** *Let* $\Omega = \{a, b, c, d, e\}$*, and let* $\Theta = \{\theta_1, \theta_2, \theta_3\}$ *be the partition of* $\Omega$ *defined by*

$$\theta_1 = \{a, b\}, \quad \theta_2 = \{c, d\}, \quad \theta_3 = \{e\}.$$

*The set* $\Theta$ *is a coarsening of* $\Omega$*, and the corresponding refining is the mapping* $\rho : 2^\Theta \to 2^\Omega$ *defined by:*

$$\begin{array}{ll} \rho(\{\theta_1\}) = \{a, b\} & \rho(\{\theta_2\}) = \{c, d\} \\ \rho(\{\theta_3\}) = \{e\} & \rho(\{\theta_1, \theta_2\}) = \{a, b, c, d\} \\ \rho(\{\theta_1, \theta_3\}) = \{a, b, e\} & \rho(\{\theta_2, \theta_3\}) = \{c, d, e\} \\ \rho(\Theta) = \Omega & \rho(\emptyset) = \emptyset \ . \end{array}$$

Defining the inverse operation, i.e., associating a subset of $\Theta$ to each subset of $\Omega$ is not so easy, because a refining $\rho : 2^\Theta \to 2^\Omega$ is not, in general, onto; as remarked by Shafer [15], there are usually subsets $A$ of $\Omega$ which are not "discerned" by $\Theta$, i.e., which are not equal to $\rho(B)$ for any $B \subseteq \Theta$. However, we may consider, for each subset $A$ of $\Omega$, the set of all $\theta \in \Theta$ whose images by $\rho$ are included in $A$, or the set of all $\theta \in \Theta$ whose images have a non empty intersection with $A$. This leads to the concepts of *inner reduction* $\underline{\theta}$ and *outer reduction* $\overline{\theta}$ introduced by Shafer [15, pages 117-118] and defined, respectively, as the functions from $2^\Omega$ to $2^\Theta$ verifying:

$$\underline{\theta}(A) = \{\theta \in \Theta | \rho(\{\theta\}) \subseteq A\} \tag{14}$$

$$\overline{\theta}(A) = \{\theta \in \Theta | \rho(\{\theta\}) \cap A \neq \emptyset\} \tag{15}$$

for all $A \subseteq \Omega$.

When an arbitrary subset $A$ of $\Omega$ is mapped to a subset of $\Theta$ via the inner or the outer reduction, and then carried back to $\Omega$ using the refining operation, some information is usually lost in the process and $A$ is not recovered. However, we get some sort of "approximation" of $A$ in the form of a subset and a superset, as shown by the following theorem.

**Theorem 1** *Let $\Omega$ be a frame of discernment, $\Theta$ a coarsening of $\Omega$, and $\rho$, $\underline{\theta}$ and $\overline{\theta}$ the associated refining, inner reduction and outer reduction, respectively. Then we have, for all subset $A$ of $\Omega$:*

$$\rho(\underline{\theta}(A)) \subseteq A \subseteq \rho(\overline{\theta}(A))$$

*Proof:* See the proof of Theorem 6.3 in [15, p.118]. □

**Example 2** *Let us again consider $\Omega$ and $\Theta$ defined in Example 1. Let $A = \{a, b, c\} \subseteq \Omega$. We have:*

$$\underline{\theta}(A) = \{\theta_1\} \qquad \overline{\theta}(A) = \{\theta_1, \theta_2\}$$

$$\rho(\{\theta_1\}) = \{a, b\} \qquad \rho(\{\theta_1, \theta_2\}) = \{a, b, c, d\}$$

*and*

$$\{a, b\} \subseteq A \subseteq \{a, b, c, d\}.$$

### Extension to bba's

The above operations can easily be extended from sets to bba's. More generally, let $\Omega_1$ and $\Omega_2$ be two finite sets, and $\varphi$ a mapping from $2^{\Omega_1}$ to $2^{\Omega_2}$. Then $\varphi$ may be extended to bba's using the following definition (in the sequel, the superscript of a bba will always indicate its domain).

**Definition 1** *Let $m^{\Omega_1}$ be a bba on $\Omega_1$, and let $m^{\Omega_2}$ be a bba on $\Omega_2$. We say that $m^{\Omega_2}$ is the image of $m^{\Omega_1}$ by $\varphi$, and we note $m^{\Omega_2} = \varphi(m^{\Omega_1})$, if*

$$m^{\Omega_2}(A) = \begin{cases} \sum_{\{B \subseteq \Omega_1, \varphi(B) = A\}} m^{\Omega_1}(B) & \text{if } \{B \subseteq \Omega_1, \varphi(B) = A\} \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

*for all $A \subseteq \Omega_2$.*

Hence, applying function $\varphi$ to $m_1^{\Omega}$ may be seen as transferring each mass $m^{\Omega_1}(B)$ to $\varphi(B)$, for all $B \subseteq \Omega_1$.

Using this general definition, we may define the *vacuous extension* in $\Omega$ of a bba $m^{\Theta}$ defined on $\Theta$ as the bba $\rho(m^{\Theta})$ (the notion of vacuous extension was introduced by Shafer [15]). Similarly, the inner and outer reductions of a bba may be defined as:

$$\underline{m}^{\Theta}(B) = \underline{\theta}(m^{\Omega}) \tag{16}$$

$$\overline{m}^\Theta(B) = \overline{\theta}(m^\Omega) \tag{17}$$

Note that $\overline{m}^\Theta$ is called the restriction of $m^\Omega$ by Shafer [15, p. 126].

The following theorem extends Theorem 1 from sets to bba's.

**Theorem 2** *Let $\Omega$ be a frame of discernment, $\Theta$ a coarsening of $\Omega$, and $\rho$, $\underline{\theta}$ and $\overline{\theta}$ the associated refining, inner reduction and outer reduction, respectively. Let $m^\Omega$ be a bba defined on $\Omega$, and let $\underline{m}^\Omega$ and $\overline{m}^\Omega$ be the bba's defined as:*

$$\underline{m}^\Omega = \rho(\underline{\theta}(m^\Omega))$$

$$\overline{m}^\Omega = \rho(\overline{\theta}(m^\Omega)) \ .$$

*Then, we have:*

$$\underline{m}^\Omega \subseteq m^\Omega \subseteq \overline{m}^\Omega \ .$$

*Proof:* We have, by construction,

$$\underline{m}^\Omega(A) = \sum_{\{B\subseteq\Omega, A=\rho(\underline{\theta}(B))\}} m^\Omega(B) \quad \forall A \subseteq \Omega \tag{18}$$

$$\overline{m}^\Omega(A) = \sum_{\{B\subseteq\Omega, A=\rho(\overline{\theta}(B))\}} m^\Omega(B) \quad \forall A \subset \Omega \tag{19}$$

From Theorem 1, we have $\rho(\underline{\theta}(B)) \subseteq B$ for all $B \subseteq \Omega$. Hence, the mass $\underline{m}^\Omega(A)$ is the sum of masses $m^\Omega(B)$ initially attached to supersets of $A$, which implies that $\underline{m}^\Omega \subseteq m^\Omega$. Similarly, $B \subseteq \rho(\overline{\theta}(B))$ for all $B \subseteq \Omega$, which implies that the mass $\overline{m}^\Omega(A)$ is the sum of masses $m^\Omega(B)$ initially attached to subsets of $A$; consequently, $m^\Omega \subseteq \overline{m}^\Omega$. □

**Example 3** *Once again, let us consider $\Omega$ and $\Theta$ defined in Example 1. Let $m^\Omega$ be the bba on $\Omega$ defined by*

$$m^\Omega(\{a\}) = 0.2 \qquad m^\Omega(\{a,b,c\}) = 0.3$$

$$m^\Omega(\{d,e\}) = 0.4 \qquad m^\Omega(\Omega) = 0.1$$

*We have:*

$$\underline{m}^\Theta(\emptyset) = 0.2 \qquad \underline{m}^\Theta(\{\theta_1\}) = 0.3$$

$$\underline{m}^{\Theta}(\{\theta_3\}) = 0.4 \qquad \underline{m}^{\Theta}(\Theta) = 0.1$$

*and*

$$\overline{m}^{\Theta}(\{\theta_1\}) = 0.2 \qquad \overline{m}^{\Theta}(\{\theta_1, \theta_2\}) = 0.3$$

$$\overline{m}^{\Theta}(\{\theta_2, \theta_3\}) = 0.4 \qquad \overline{m}^{\Theta}(\Theta) = 0.1$$

*Let* $\underline{m}^{\Omega} = \rho(\underline{m}^{\Theta})$ *and* $\overline{m}^{\Omega} = \rho(\overline{m}^{\Theta})$ *the vacuous extensions of* $\underline{m}^{\Theta}$ *and* $\overline{m}^{\Theta}$*, respectively. We have*

$$\underline{m}^{\Omega}(\emptyset) = 0.2 \qquad \underline{m}^{\Omega}(\{a, b\}) = 0.3$$

$$\underline{m}^{\Omega}(\{e\}) = 0.4 \qquad \underline{m}^{\Omega}(\Omega) = 0.1$$

*and*

$$\overline{m}^{\Omega}(\{a, b\}) = 0.2 \qquad \overline{m}^{\Omega}(\{a, b, c, d\}) = 0.3$$

$$\overline{m}^{\Omega}(\{c, d, e\}) = 0.4 \qquad \overline{m}^{\Omega}(\Omega) = 0.1$$

*It may be checked that* $\underline{m}^{\Omega} \subseteq m^{\Omega} \subseteq \overline{m}^{\Omega}$*.*

As shown by the above theorem, the definition of a coarsening of the frame $\Omega$ allows to define strong inner and outer approximations of any bba on $\Omega$. This principle will be exploited in Section 3 to define a method for approximating the combination of several bba's. To implement such a method, we need a representation of bba's allowing to carry a bba easily from one frame to another. Such a representation is introduced in the next section.

## 2.4 Matrix representation of bba's

A very simple construction of $\underline{m}^{\Omega}$ and $\overline{m}^{\Omega}$ for a given coarsening $\Theta$ can be obtained using the following representation. Let us assume that the frame $\Omega = \{\omega_1, \ldots \omega_n\}$ has $n$ elements, and the bba $m^{\Omega}$ under consideration has $p$ focal elements: $\mathcal{F}(m^{\Omega}) = \{A_1, \ldots, A_p\}$. One can represent the bba $m^{\Omega}$ by a pair $(\mathbf{m}^{\Omega}, \mathbf{F}^{\Omega})$ where $\mathbf{m}^{\Omega}$ is the $p$-dimensional column vector of masses:

$$\mathbf{m}^{\Omega} = \begin{bmatrix} m^{\Omega}(A_1) \\ \vdots \\ m^{\Omega}(A_p) \end{bmatrix}$$

and $\mathbf{F}^\Omega$ is a $p \times n$ binary matrix such that

$$\mathbf{F}^\Omega_{ij} = A_i(\omega_j) = \begin{cases} 1, & \text{if } \omega_j \in A_i \\ 0, & \text{otherwise.} \end{cases}$$

where $A_i(\cdot)$ denotes the indicator function of focal element $A_i$.

This representation is similar to an (objects $\times$ attributes) binary data matrix as commonly considered in data analysis. Here, each focal element can be seen as an object, and each element of the frame corresponds to a binary attribute. Each object $A_i$ has a weight $m^\Omega(A_i)$. Since a coarsening is inherently equivalent to a partition of $\Omega$, finding a suitable coarsening is actually a problem of classifying the columns of data matrix $\mathbf{F}^\Omega$, which is a classical clustering problem (see, e.g. [7]). Note that, in contrast, the clustering approximation method introduced by Denœux [2] is based on the classification of the rows of $\mathbf{F}^\Omega$.

To see how the matrix representations of $\underline{m}^\Theta$, $\overline{m}^\Theta$, $\underline{m}^\Omega$, $\overline{m}^\Omega$ can be obtained, let us denote by $P = \{I_1, \ldots, I_c\}$ the partition of $N_n = \{1, \ldots, n\}$ corresponding to the coarsening $\Theta = \{\theta_1, \ldots, \theta_c\}$, i.e.,

$$\theta_r = \{\omega_j, j \in I_r\} \qquad r = 1, \ldots, c.$$

Let $(\underline{\mathbf{m}}^\Theta, \underline{\mathbf{F}}^\Theta)$ denote the matrix representation of $\underline{m}^\Theta$. Matrix $\underline{\mathbf{F}}^\Theta$ may be obtained from $\mathbf{F}^\Omega$ by merging the columns $\mathbf{F}^\Omega_{\cdot j}$ for $j \in I_r$, replacing them by their minimum:

$$\underline{\mathbf{F}}^\Theta_{i,r} = \min_{j \in I_r} \mathbf{F}^\Omega_{i,j} \quad \forall i, r \tag{20}$$

and leaving the mass vector unchanged: $\underline{\mathbf{m}}^\Theta = \mathbf{m}^\Omega$. The justification for this is that the focal elements of $\underline{m}^\Theta$ are the sets $\underline{\theta}(A_i)$, and $\theta_r \in \underline{\theta}(A_i)$ iff $\rho(\theta_r) \subseteq A_i$, where $\rho$ is the refining associated to $\Theta$. Hence, the column corresponding to $\theta_r$ in $\underline{\mathbf{F}}^\Theta$ has a 1 in row $i$, if and only if $\mathbf{F}^\Omega_{i,j} = 1$ for all $j \in I_r$.

Similarly, if $(\overline{\mathbf{m}}^\Theta, \overline{\mathbf{F}}^\Theta)$ denotes the matrix representation of $\overline{m}^\Theta$, we have

$$\overline{\mathbf{F}}^\Theta_{i,r} = \max_{j \in I_r} \mathbf{F}^\Omega_{i,j} \quad \forall i, r \tag{21}$$

and $\overline{\mathbf{m}}^\Theta = \mathbf{m}^\Omega$.

The matrix representations of $\underline{m}^\Omega$ and $\overline{m}^\Omega$, the vacuous extensions of $\underline{m}^\Theta$ and $\overline{m}^\Theta$, are then obtained by duplicating $|I_r|$ times each column $\underline{\mathbf{F}}^\Theta_{\cdot r}$ and $\overline{\mathbf{F}}^\Theta_{\cdot r}$, respectively:

$$\underline{\mathbf{F}}^\Omega_{i,j} = \underline{\mathbf{F}}^\Theta_{i,r} \quad \forall j \in I_r \tag{22}$$

$$\overline{\mathbf{F}}^{\Omega}_{i,j} = \overline{\mathbf{F}}^{\Theta}_{i,r} \quad \forall j \in I_r \ . \tag{23}$$

Again, the mass vector is unchanged: $\underline{\mathbf{m}}^{\Omega} = \overline{\mathbf{m}}^{\Omega} = \mathbf{m}^{\Omega}$.

**Remark 2** *Several rows of* $\underline{\mathbf{F}}^{\Omega}$ *or* $\overline{\mathbf{F}}^{\Omega}$ *computed by the above method may be identical. In this case, similar raws of the binary matrix of focal elements have to be merged and the masses have to be added, so that the row dimension becomes equal to the number of focal elements.*

**Example 4** *Table 1 shows how the matrix representation can be used to recover the results of Example 3.*

Table 1: Calculation of inner (left) and outer (right) approximations using the matrix representation

| | $a$ | $b$ | $c$ | $d$ | $e$ | |
|---|---|---|---|---|---|---|
| $m^{\Omega}$ | 1 | 0 | 0 | 0 | 0 | 0.2 |
| | 1 | 1 | 1 | 0 | 0 | 0.3 |
| | 0 | 0 | 0 | 1 | 1 | 0.4 |
| | 1 | 1 | 1 | 1 | 1 | 0.1 |
| | $\theta_1$ | | $\theta_2$ | | $\theta_3$ | |
| $\underline{m}^{\Theta}$ | 0 | | 0 | | 0 | 0.2 |
| | 1 | | 0 | | 0 | 0.3 |
| | 0 | | 0 | | 1 | 0.4 |
| | 1 | | 1 | | 1 | 0.1 |
| | $a$ | $b$ | $c$ | $d$ | $e$ | |
| $\underline{m}^{\Omega}$ | 0 | 0 | 0 | 0 | 0 | 0.2 |
| | 1 | 1 | 0 | 0 | 0 | 0.3 |
| | 0 | 0 | 0 | 0 | 1 | 0.4 |
| | 1 | 1 | 1 | 1 | 1 | 0.1 |

| | $a$ | $b$ | $c$ | $d$ | $e$ | |
|---|---|---|---|---|---|---|
| $m^{\Omega}$ | 1 | 0 | 0 | 0 | 0 | 0.2 |
| | 1 | 1 | 1 | 0 | 0 | 0.3 |
| | 0 | 0 | 0 | 1 | 1 | 0.4 |
| | 1 | 1 | 1 | 1 | 1 | 0.1 |
| | $\theta_1$ | | $\theta_2$ | | $\theta_3$ | |
| $\overline{m}^{\Theta}$ | 1 | | 0 | | 0 | 0.2 |
| | 1 | | 1 | | 0 | 0.3 |
| | 0 | | 1 | | 1 | 0.4 |
| | 1 | | 1 | | 1 | 0.1 |
| | $a$ | $b$ | $c$ | $d$ | $e$ | |
| $\overline{m}^{\Omega}$ | 1 | 1 | 0 | 0 | 0 | 0.2 |
| | 1 | 1 | 1 | 1 | 0 | 0.3 |
| | 0 | 0 | 1 | 1 | 1 | 0.4 |
| | 1 | 1 | 1 | 1 | 1 | 0.1 |

# 3 Coarsening Approximations of Belief Functions

## 3.1 Principle

In this section, we propose a new heuristic method for constructing strong inner and outer approximations of the combination of several bba's. Our method consists in performing the combination in a coarsening $\Theta$ of the initial frame $\Omega$, using the FMT algorithm (which allows to reduce the complexity from $2^{|\Omega|}$ to $2^{|\Theta|}$), and carrying back the result to $\Omega$ using the vacuous extension. This approach is summarized in Figure 1, for the case of the inner approximation of the conjunctive sum of 2 bba's. It can be easily generalized to the inner and outer approximation of the conjunctive or disjunctive combination of several bba's.

Before addressing the issue of the choice of the coarsening $\Theta$ in which to perform the combination, we need to prove that the above scheme indeed leads to strong inner and outer approximations. This is the subject of the following theorem:

**Theorem 3** *Let $m_1^\Omega, \ldots, m_K^\Omega$ be $K$ bba's defined on a frame $\Omega$, $\underline{m}_1^\Theta, \ldots, \underline{m}_K^\Theta$ (resp, $\overline{m}_1^\Theta, \ldots, \overline{m}_K^\Theta$) their inner (resp., outer) reductions in a coarsening $\Theta$ of $\Omega$. Let $\underline{m}^\Theta$ and $\overline{m}^\Theta$ be defined as:*

$$\underline{m}^\Theta = \underline{m}_1^\Theta \odot \ldots \odot \underline{m}_K^\Theta$$

$$\overline{m}^\Theta = \overline{m}_1^\Theta \odot \ldots \odot \overline{m}_K^\Theta,$$

*where $\odot \in \{\textcircled{\cap}, \textcircled{\cup}\}$ denotes the conjunctive or the disjunctive sum. Let $\underline{m}^\Omega$ and $\overline{m}^\Omega$ be the vacuous extensions of $\underline{m}^\Theta$ and $\overline{m}^\Theta$ in $\Omega$. Then:*

$$\underline{m}^\Omega \subseteq m^\Omega \subseteq \overline{m}^\Omega$$

*where $m^\Omega = m_1^\Omega \odot \ldots \odot m_K^\Omega$.*

*Proof:* As shown by Shafer [15, p.166], we have, for all subsets $B_1, \ldots, B_K$ of $\Theta$,

$$\rho(\bigcap_{k=1}^{K} B_k) = \bigcap_{k=1}^{K} \rho(B_k)$$

$$\rho(\bigcup_{k=1}^{K} B_k) = \bigcup_{k=1}^{K} \rho(B_k)\ .$$

$$
\begin{array}{ccc}
m_1^{\Omega}, m_2^{\Omega} & \xrightarrow{\theta} & \underline{m}_1^{\Theta}, \underline{m}_2^{\Theta} \\
 & & \downarrow \text{FMT} \\
 & & \underline{q}^{\Theta} = \underline{q}_1^{\Theta} \cdot \underline{q}_2^{\Theta} \\
 & & \downarrow \text{FMT} \\
\underline{m}^{\Omega} & \xleftarrow{\rho} & \underline{m}^{\Theta} = \underline{m}_1^{\Theta} \bigcirc\hspace{-0.75em}\cap\; \underline{m}_2^{\Theta}
\end{array}
$$

Figure 1: Inner approximation of the conjunctive sum of 2 bba's.

As a consequence, the vacuous extension commutes with the disjunctive or conjunctive combination, i.e.,we have:

$$\rho(\underline{m}_1^{\Theta} \odot \ldots \odot \underline{m}_K^{\Theta}) = \rho(\underline{m}_1^{\Theta}) \odot \ldots \odot \rho(\underline{m}_K^{\Theta}).$$

It follows from Theorem 2 that

$$\rho(\underline{m}_i^{\Theta}) \subseteq m_i^{\Omega}, \quad i = 1, \ldots, K.$$

Proposition 2 thus entails that

$$\rho(\underline{m}_1^{\Theta}) \odot \ldots \odot \rho(\underline{m}_K^{\Theta}) \subseteq m_1^{\Omega} \odot \ldots \odot m_K^{\Omega},$$

that is, $\underline{m}^{\Omega} \subseteq m^{\Omega}$. A similar argument shows that $m^{\Omega} \subseteq \overline{m}^{\Omega}$. □

## 3.2 The method

### Measuring the information loss

As shown by Theorem 3, the combination of $K$ bba's $m_1^{\Omega}, \ldots, m_K^{\Omega}$ defined on a frame $\Omega$ can be approximated by combining the inner and outer reductions of the $K$ bba's in a coarsening $\Theta$. The quality of the approximation will obviously depend on the coarsening chosen. When carrying the bba's to $\Theta$, combining them, and carrying back the result to $\Omega$, some information is lost, and this loss of information should naturally be minimized.

One of the first problems to be solved is the measurement of the information contained in a piece of evidence. This is a complex problem for which many approaches have been proposed [12, 13, 10], without any single measure undoubtedly emerging as the best one. One of these measures is the *generalized cardinality* [4, 2], which is defined for a bba $m$:

$$|m| = \sum_{i=1}^{p} m(A_i)|A_i|, \tag{24}$$

where $A_i$, $i = 1, \ldots, p$ are the focal elements of $m$. The bba $m$ is all the more imprecise, or unspecific (and contains all the less information) that $|m|$ is large. It follows from the definition of strong inclusion that

$$m \subseteq m' \Rightarrow |m| \leq |m'|. \tag{25}$$

**Remark 3** *Let $A_i(\cdot)$ denote the indicator function of focal element $A_i$. We have*

$$\begin{aligned} |m| &= \sum_{i=1}^{p} m(A_i) \sum_{\omega \in \Omega} A_i(\omega) \\ &= \sum_{\omega \in \Omega} \sum_{i=1}^{p} m(A_i) A_i(\omega) \\ &= \sum_{\omega \in \Omega} pl(\{\omega\}), \end{aligned}$$

*which provides another interpretation of $|m|$ as a measure of uncertainty*[1].

Assume that we have a bba $m^\Omega$ which we want to represent in a coarsening $\Theta$ of size $|\Theta| = c$ using the outer reduction operation. Because $m^\Omega$ and $\overline{m}^\Theta$ are in different frames, it is not relevant to compare their cardinality. However, it results from Theorem 2 that $|m^\Omega| \leq |\rho(\overline{m}^\Theta)|$. Hence, the loss of information induced by the mapping to a coarsening may be measured by the quantity

$$\Delta(\rho(\overline{m}^\Theta), m^\Omega) = |\rho(\overline{m}^\Theta)| - |m^\Omega|.$$

If this quantity is zero, it means that no information has been lost when carrying $m^\Omega$ to $\Theta$, since $m^\Omega$ can be recovered by the inverse process.

**Remark 4** *Let $\Theta = \{\theta_1, \ldots, \theta_c\}$, and $\Omega_i = \rho(\{\theta_i\})$. Hence $\{\Omega_1, \ldots, \Omega_c\}$ is the partition of $\Omega$ induced by the coarsening. We have*

$$|\rho(\overline{m}^\Theta)| = \sum_{\omega \in \Omega} \overline{pl}^\Omega(\{\omega\}) = \sum_{i=1}^{c} |\Omega_i| \overline{pl}^\Omega(\Omega_i).$$

[1] This remark, as well as Remarks 4 and 5, were suggested to the authors by Philippe Smets.

*But we may observe that* $\overline{pl}(\Omega_i) = pl(\Omega_i)$ *for all* $i \in \{1, \ldots, c\}$*. Hence,*

$$|\rho(\overline{m}^{\Theta})| = \sum_{i=1}^{c} |\Omega_i| pl^{\Omega}(\Omega_i).$$

Similarly, in the case of the inner reduction, we have $|\rho(\overline{m}^{\Theta})| \leq |m^{\Omega}|$. In that case, $|\rho(\overline{m}^{\Theta})|$ is thus more precise than $m^{\Omega}$, but this gain of precision corresponds to "irrelevant", or "meaningless" information, and, consequently, it should be minimized. The gain of irrelevant information is measured by

$$\Delta(m^{\Omega}, \rho(\underline{m}^{\Theta})) = |m^{\Omega}| - |\rho(\underline{m}^{\Theta})|.$$

**Example 5** *Let us go back to Example 3. We have*

$$|m^{\Omega}| = 2.4 \qquad |\rho(\underline{m}^{\Theta})| = |\underline{m}^{\Omega}| = 1.1 \qquad |\rho(\overline{m}^{\Theta})| = |\overline{m}^{\Omega}| = 3.3$$

*Now, let us consider another coarsening* $\Theta' = \{\{a, c\}, \{b\}, \{d, e\}\}$*, and let* $\rho'$ *denote the corresponding refining. It can easily be checked, using the technique presented in Section 2.4, that we have*

$$\rho'(\underline{m}^{\Theta'})(\emptyset) = 0.2$$

$$\rho'(\underline{m}^{\Theta'})(\{a, b, c\}) = 0.3$$

$$\rho'(\underline{m}^{\Theta'})(\{d, e\}) = 0.4$$

$$\rho'(\underline{m}^{\Theta'})(\Omega) = 0.1$$

*and*

$$\rho'(\overline{m}^{\Theta'})(\{a, c\}) = 0.2$$

$$\rho'(\overline{m}^{\Theta'})(\{a, b, c\}) = 0.3$$

$$\rho'(\overline{m}^{\Theta'})(\{d, e\}) = 0.4$$

$$\rho'(\overline{m}^{\Theta'})(\Omega) = 0.1$$

*Hence, we have* $|\rho'(\underline{m}^{\Theta'})| = 2.2$ *and* $|\rho'(\overline{m}^{\Theta'})| = 2.6$*. We notice that*

$$\Delta(\rho(\overline{m}^{\Theta'}), m^{\Omega}) < \Delta(\rho(\overline{m}^{\Theta}), m^{\Omega})$$

*and*

$$\Delta(m^{\Omega}, \rho(\underline{m}^{\Theta'})) < \Delta(m^{\Omega}, \rho(\underline{m}^{\Theta})).$$

*Consequently,* $\Theta'$ *can be considered to induce better inner and outer approximations than* $\Theta$*.*

**Optimal coarsenings**

Let us assume that we want to approximate the combination of $K$ bba's $m_1^\Omega, \ldots, m_K^\Omega$, using a coarsening $\Theta$ of $\Omega$. The quality of the inner approximations of $m_1^\Omega, \ldots, m_K^\Omega$ may be measured globally by

$$\sum_{k=1}^{K} \Delta(m_k^\Omega, \rho(\underline{m}_k^\Theta)).$$

Let us denote by $\mathcal{K}_c$ the set of all coarsenings of $\Omega$ obtained by partitioning $\Omega$ in $c$ classes ($c < n$). The coarsening $\underline{\Theta}_c$ yielding the "best" (least specific) inner approximation of the $K$ bba's $m_1^\Omega, \ldots, m_K^\Omega$ may then be defined as:

$$\underline{\Theta}_c = \arg \min_{\Theta \in \mathcal{K}_c} \sum_{k=1}^{K} \Delta(m_k^\Omega, \rho(\underline{m}_k^\Theta)) \tag{26}$$

Similarly, the coarsening $\overline{\Theta}_c$ yielding the best (most specific) outer approximation shall be defined as

$$\overline{\Theta}_c = \arg \min_{\Theta \in \mathcal{K}_c} \sum_{k=1}^{K} \Delta(\rho(\overline{m}_k^\Theta), m_k^\Omega) \tag{27}$$

**Hierarchical clustering approach**

Given the potentially huge number of partitions of a set of $n$ elements into $c$ subsets, solving problems (26) and (27) by exhaustive enumeration is not computationally feasable. In the absence of an efficient algorithm for finding the global optimum for these problems, a heuristic procedure has to be used. One such approach is hierarchical clustering [7], a well-known method for constructing a sequence of nested partitions of a given set. In our case, this approach will consist in sequentially aggregating pairs of elements of $\Omega$ until the desired size of the coarsened frame of discernment is reached. At each step, the two elements whose aggregation results in the best value of the criterion will be selected.

More precisely, let $(\mathbf{m}_k^\Omega, \mathbf{F}_k^\Omega)$, $k = 1, \ldots, K$ denote the matrix representations of the $K$ bba's $m_1^\Omega, \ldots, m_K^\Omega$ to be combined. $\mathbf{m}_k^\Omega$ is a column vector of length $p_k$ (the number of focal elements of $m_k^\Omega$), and $\mathbf{F}_k^\Omega$ is a matrix of size $p_k \times n$. Let $\mathbf{F}^\Omega$ and $\mathbf{m}^\Omega$

be defined as

$$\mathbf{F}^\Omega = \begin{bmatrix} \mathbf{F}_1^\Omega \\ \vdots \\ \mathbf{F}_K^\Omega \end{bmatrix} \qquad \mathbf{m}^\Omega = \begin{bmatrix} \mathbf{m}_1^\Omega \\ \vdots \\ \mathbf{m}_K^\Omega \end{bmatrix}$$

Hence, $\mathbf{m}^\Omega$ is a column vector of size $p = \sum_{k=1}^K p_k$, and $\mathbf{F}^\Omega$ is a matrix of size $p \times n$. The pair $(\mathbf{m}^\Omega, \mathbf{F}^\Omega)$ will be called the (joint) matrix representation of $m_1^\Omega, \ldots, m_K^\Omega$.

Suppose that we are looking for the coarsening with $n-1$ elements corresponding to the "best" inner approximation. As shown in Section 2.4, the aggregation of elements $\omega_j$ and $\omega_l$ of the frame corresponds to the fusion of columns $j$ and $l$ of $\mathbf{F}^\Omega$ using the minimum operator. In this operation, the number of 1's in each row $i$ of matrix $\mathbf{F}^\Omega$ is decreased by one if either $\omega_j \in A_i$ and $\omega_l \notin A_i$, or $\omega_l \in A_i$ and $\omega_j \notin A_i$. Hence, the decrease of cardinality is

$$\delta(j,l) = \sum_{k=1}^{K} \Delta(m_k^\Omega, \rho_{jl}(\underline{m}_k^{\Theta_{jl}})) = \sum_{i=1}^{p} \mathbf{m}_i^\Omega |\mathbf{F}_{ij}^\Omega - \mathbf{F}_{il}^\Omega| \tag{28}$$

where $\mathbf{F}_{ij}^\Omega$ and $\mathbf{F}_{il}^\Omega$ are, respectively, the elements $(i,j)$ and $(i,l)$ of matrix $\mathbf{F}^\Omega$, $\Theta_{jl}$ is the coarsening of $\Omega$ obtained by merging $\omega_j$ and $\omega_l$, and $\rho_{jl}$ is the corresponding refining. Note that $\delta(\omega_j, \omega_l)$ can be interpreted as a degree of dissimilarity [7] between $\omega_j$ and $\omega_l$.

By computing $\delta(j,l)$ for each ordered pair of columns, one obtains a dissimilarity matrix $D = \delta(j,l), j,l \in \{1,\ldots,n\}$. The pair $(\omega_{j^*}, \omega_{l^*})$ such that

$$\delta(j^*, l^*) = \min_{j,l} \delta(j,l)$$

is selected, and a new matrix $\underline{\mathbf{F}}^{\Theta_{j^*l^*}}$ with $n-1$ columns is constructed by aggregating columns $j^*$ and $l^*$ of $\mathbf{F}^\Omega$. The dissimilarity matrix $D$ is then updated, and the whole process is iterated until a frame $\Theta$ of desired size is obtained. The output of the algorithm is then the joint matrix representation $(\underline{\mathbf{m}}^\Theta, \underline{\mathbf{F}}^\Theta)$ of the K bba's $\underline{m}_1^\Theta, \ldots, \underline{m}_K^\Theta$.

The computation of outer approximations can be performed in exactly the same way, except that the minimum operator is replaced by the maximum operator. After aggregating columns $j$ and $l$ of matrix $\mathbf{F}^\Omega$, the number of 1's in each row $i$ of matrix $\mathbf{F}^\Omega$ is now increased by one if either $\omega_j \in A_i$ and $\omega_l \notin A_i$, or $\omega_l \in A_i$ and $\omega_j \notin A_i$.

Hence, the increase of cardinality is

$$\sum_{k=1}^{K} \Delta(\rho_{jl}(\overline{m}_k^{\Theta_{jl}}), m_k^{\Omega}) = \sum_{i=1}^{p} \mathbf{m}_i^{\Omega} |\mathbf{F}_{ij}^{\Omega} - \mathbf{F}_{il}^{\Omega}| = \delta(j,l) \tag{29}$$

We thus arrive at the same dissimilarity measure as in the previous case. However, after a few steps, the resulting coarsening will generally be different, because a different operator is used for fusing two columns at each step.

Note that the above algorithm is basically the classical hierarchical clustering algorithm applied to the binary matrix of focal elements. Hence, the time needed to compute an inner or outer coarsening approximation by this method is proportional to $n^3$.

**Remark 5** *As a consequence of Remark 4, we may observe that*

$$\begin{aligned} \Delta(\rho_{jl}(\overline{m}_k^{\Theta_{jl}}), m_k^{\Omega}) &= |\rho_{jl}(\overline{m}_k^{\Theta_{jl}})| - |m_k^{\Omega}| \\ &= 2pl_k^{\Omega}(\{\omega_j, \omega_l\}) + \sum_{i \neq j, i \neq l} pl_k^{\Omega}(\{\omega_i\}) - \sum_{\omega \in \Omega} pl_k^{\Omega}(\{\omega\}) \\ &= 2pl_k^{\Omega}(\{\omega_j, \omega_l\}) - pl_k^{\Omega}(\{\omega_j\}) - pl_k^{\Omega}(\{\omega_l\}). \end{aligned}$$

*This leads to the following simple expression for $\delta(j,l)$:*

$$\delta(j,l) = 2\sum_{k=1}^{K} pl_k^{\Omega}(\{\omega_j, \omega_l\}) - \sum_{k=1}^{K} pl_k^{\Omega}(\{\omega_j\}) - \sum_{k=1}^{K} pl_k^{\Omega}(\{\omega_l\}). \tag{30}$$

**Inner and Outer approximations of the Combined Belief Function**

Given $K$ bba's $\underline{m}_1^{\Theta}, \ldots, \underline{m}_K^{\Theta}$ and $\overline{m}_1^{\Theta}, \ldots, \overline{m}_K^{\Theta}$ defined over the common coarsened frame $\Theta$ of $\Omega$, we finally proceed as follows to determine strong inner and outer approximations of their combination:

1. Use the FMT algorithm to convert these approximated bba's into their related inner and outer commonality or implicability functions.

2. Multiply the commonality or implicability functions over the coarsened frame $\Theta$. In the case of inner approximation we have: $\underline{q}^{\Theta} = \prod_{k=1}^{K} \underline{q}_k^{\Theta}$ and $\underline{b}^{\Theta} = \prod_{k=1}^{K} \underline{b}_k^{\Theta}$, and similarly for the outer approximations $\overline{q}^{\Theta}$ and $\overline{b}^{\Theta}$.

3. Convert back these approximated combined commonality or implicability functions into the associated mass functions $\underline{m}^\Theta$ and $\overline{m}^\Theta$ using the FMT algorithm.

4. Using the vacuous extension, generate the inner and outer approximations $\underline{m}^\Omega$ and $\overline{m}^\Omega$.

In the above procedure, the combination of the $K$ bba's is performed in the coarsened frame $\Theta$. Consequently, the time needed to compute the combination by this method is proportional to $Kc^2 2^c$ with $c = |\Theta|$, which is potentially much smaller than the time needed to do the combination in $\Omega$. Taking into account the construction of the coarsenings (which takes time proportional to $n^3$), the time needed by the overall procedure is thus roughly proportional to $\max(n^3, Kc^2 2^c)$.

### 3.3 Example

Let us consider the following bba's on $\Omega = \{a, b, c, d, e\}$:

$$m_1^\Omega(\{a,b\}) = 0.5 \qquad m_1^\Omega(\{a,c,d\}) = 0.3$$

$$m_1^\Omega(\{c,d\}) = 0.1 \qquad m_1^\Omega(\{d,e\}) = 0.1$$

and

$$m_2^\Omega(\{a,e\}) = 0.4 \qquad m_2^\Omega(\{b,d\}) = 0.3$$

$$m_2^\Omega(\{a,b,e\}) = 0.2 \qquad m_2^\Omega(\{a\}) = 0.1$$

The joint matrix representation of $m_1^\Omega$ and $m_2^\Omega$ is shown in Table 2.

Assume that we wish to compute an outer approximation of $m_1^\Omega \circledcirc m_2^\Omega$ using a coarsening $\Theta$ of size $c = 3$. This is achieved in two steps using the hierarchical clustering approach described above.

**Step 1:** The symmetric dissimilarity matrix $D = \delta(j, l)$ is computed (see Table 3). As the aggregation of $c$ and $d$ yields the smallest cardinality increase, this pair is selected. The resulting coarsening is $\Theta_{cd} = \{\{a\}, \{b\}, \{c, d\}, \{e\}\}$. The joint matrix representation of $\overline{m}_1^{\Theta_{cd}}$ and $\overline{m}_2^{\Theta_{cd}}$, obtained by aggregating columns $c$ and $d$ of Table 2 using the maximum operator, is shown in Table 4.

Table 2: Joint matrix representation of 2 bba's $m_1^\Omega$ and $m_2^\Omega$

| | $a$ | $b$ | $c$ | $d$ | $e$ | |
|---|---|---|---|---|---|---|
| $m_1^\Omega$ | 1 | 1 | 0 | 0 | 0 | 0.5 |
| | 1 | 0 | 1 | 1 | 0 | 0.3 |
| | 0 | 0 | 1 | 1 | 0 | 0.1 |
| | 0 | 0 | 0 | 1 | 1 | 0.1 |
| $m_2^\Omega$ | 1 | 0 | 0 | 0 | 1 | 0.4 |
| | 0 | 1 | 0 | 1 | 0 | 0.3 |
| | 1 | 1 | 0 | 0 | 1 | 0.2 |
| | 1 | 0 | 0 | 0 | 0 | 0.1 |

Table 3: Dissimilarity matrix at step 1

| | $a$ | $b$ | $c$ | $d$ | $e$ |
|---|---|---|---|---|---|
| $a$ | - | | | | |
| $b$ | 1.1 | - | | | |
| $c$ | 1.3 | 1.4 | - | | |
| $d$ | 1.7 | 1.2 | 0.4 | - | |
| $e$ | 1.0 | 1.3 | 1.1 | 1.3 | - |

**Step 2** The dissimilarity matrix is updated, leading to the result shown in Table 5. This time, the merging of $a$ and $e$ yields the smallest cardinality increase. The resulting coarsening is $\Theta = \{\{a, e\}, \{b\}, \{c, d\}\}$. The joint matrix representation of $\overline{m}_1^\Theta$ and $\overline{m}_2^\Theta$ is shown in Table 6.

Finally, Table 7 shows the conjunctive sum of $\overline{m}_1^\Theta$ and $\overline{m}_2^\Theta$, and its vacuous extension in $\Omega$. It can be checked that $\rho(\overline{m}_1^\Theta \textcircled{\cap} \overline{m}_2^\Theta)$ is a strong outer approximation of $m_1^\Omega \textcircled{\cap} m_2^\Omega$ shown in Table 8.

Table 4: Joint matrix representation of 2 bba's $\overline{m}_1^{\Theta_{cd}}$ and $\overline{m}_2^{\Theta_{cd}}$

| | $a$ | $b$ | $\{c,d\}$ | $e$ | |
|---|---|---|---|---|---|
| $\overline{m}_1^{\Theta_{cd}}$ | 1 | 1 | 0 | 0 | 0.5 |
| | 1 | 0 | 1 | 0 | 0.3 |
| | 0 | 0 | 1 | 0 | 0.1 |
| | 0 | 0 | 1 | 1 | 0.1 |
| $\overline{m}_2^{\Theta_{cd}}$ | 1 | 0 | 0 | 1 | 0.4 |
| | 0 | 1 | 1 | 0 | 0.3 |
| | 1 | 1 | 0 | 1 | 0.2 |
| | 1 | 0 | 0 | 0 | 0.1 |

Table 5: Dissimilarity matrix at step 2

| | $a$ | $b$ | $\{c,d\}$ | $e$ |
|---|---|---|---|---|
| $a$ | - | | | |
| $b$ | 1.1 | - | | |
| $\{c,d\}$ | 1.7 | 1.2 | - | |
| $e$ | 1.0 | 1.3 | 1.3 | - |

Table 6: Joint matrix representation of 2 bba's $\overline{m}_1^{\Theta}$ and $\overline{m}_2^{\Theta}$

| | $\{a,e\}$ | $b$ | $\{c,d\}$ | |
|---|---|---|---|---|
| $\overline{m}_1^{\Theta}$ | 1 | 1 | 0 | 0.5 |
| | 1 | 0 | 1 | 0.3 |
| | 0 | 0 | 1 | 0.1 |
| | 1 | 0 | 1 | 0.1 |
| $\overline{m}_2^{\Theta}$ | 1 | 0 | 0 | 0.4 |
| | 0 | 1 | 1 | 0.3 |
| | 1 | 1 | 0 | 0.2 |
| | 1 | 0 | 0 | 0.1 |

Table 7: Conjunctive sum of $\overline{m}_1^\Theta$ and $\overline{m}_2^\Theta$, and its vacuous extension in $\Omega$

| | $\{a,e\}$ | $b$ | $\{c,d\}$ | |
|---|---|---|---|---|
| $\overline{m}_1^\Theta \textcircled{\cap} \overline{m}_2^\Theta$ | 0 | 0 | 0 | 0.07 |
| | 1 | 0 | 0 | 0.53 |
| | 0 | 1 | 0 | 0.15 |
| | 1 | 1 | 0 | 0.10 |
| | 0 | 0 | 1 | 0.15 |

| | $a$ | $b$ | $c$ | $d$ | $e$ | |
|---|---|---|---|---|---|---|
| $\rho(\overline{m}_1^\Theta \textcircled{\cap} \overline{m}_2^\Theta)$ | 0 | 0 | 0 | 0 | 0 | 0.07 |
| | 1 | 0 | 0 | 0 | 1 | 0.53 |
| | 0 | 1 | 0 | 0 | 0 | 0.15 |
| | 1 | 1 | 0 | 0 | 1 | 0.10 |
| | 0 | 0 | 1 | 1 | 0 | 0.15 |

Table 8: Conjunctive sum of $m_1^\Omega$ and $m_2^\Omega$

| | $a$ | $b$ | $c$ | $d$ | $e$ | |
|---|---|---|---|---|---|---|
| $m_1^\Omega \textcircled{\cap} m_2^\Omega$ | 0 | 0 | 0 | 0 | 0 | 0.08 |
| | 1 | 0 | 0 | 0 | 0 | 0.46 |
| | 0 | 1 | 0 | 0 | 0 | 0.15 |
| | 1 | 1 | 0 | 0 | 0 | 0.10 |
| | 0 | 0 | 0 | 1 | 0 | 0.15 |
| | 0 | 0 | 0 | 0 | 1 | 0.06 |

# 4 Simulations

As an example, we simulated the conjunctive combination of 3 bba's on a frame $\Omega$ of size $n = |\Omega| = 30$, with 500 focal elements each. The focal elements were generated randomly in such a way that element $\omega_i$ of the frame has probability $(i/(n+1))^2$ to belong to each focal element. In this way, we simulate the realistic situation in which some single hypotheses are more plausible than others. The masses were assigned to focal elements as proposed by Tessem [19]: the mass given to the first one was taken from a uniform distribution on $[0, 1]$, then a random fraction of the rest was given to the second one, etc. The remaining part of the unit mass was finally allocated to the last focal element. The conjunctive sum of the 3 bba's was approximated using the method described above, using coarsenings of various sizes.

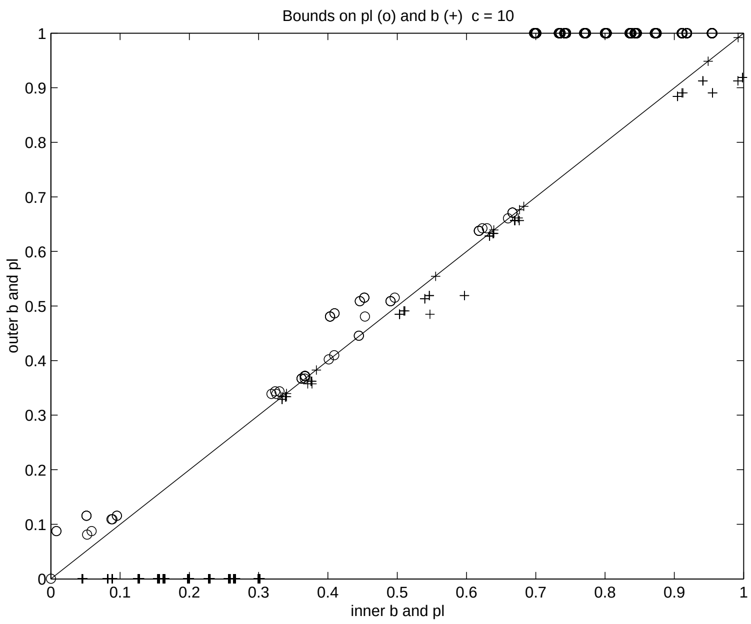


Figure 2: Inner vs. outer approximations of functions $b$ and pl (resulting from the conjunctive combination of 3 bba's) for $N = 1000$ randomly generated subsets of $\Omega$, with $|\Omega = 30|$, using a coarsening of size $c = |\Theta| = 10$.

Examples of the results obtained for one particular trial are shown in Figures 2 (for $c = 10$) and 3 (for $c = 15$). In these figures, the plausibilities $\underline{\text{pl}}^{\Omega}(A)$ and implicabilities and $\underline{b}^{\Omega}(A)$ are plotted on the $x$ axis against $\overline{\text{pl}}^{\Omega}(A)$ and $\overline{b}^{\Omega}(A)$, for $N = 1000$ randomly selected subsets of $\Omega$. As expected, we observe that $\underline{\text{pl}}^{\Omega}(A) \leq \overline{\text{pl}}^{\Omega}(A)$ for all $A \subseteq \Omega$ (as

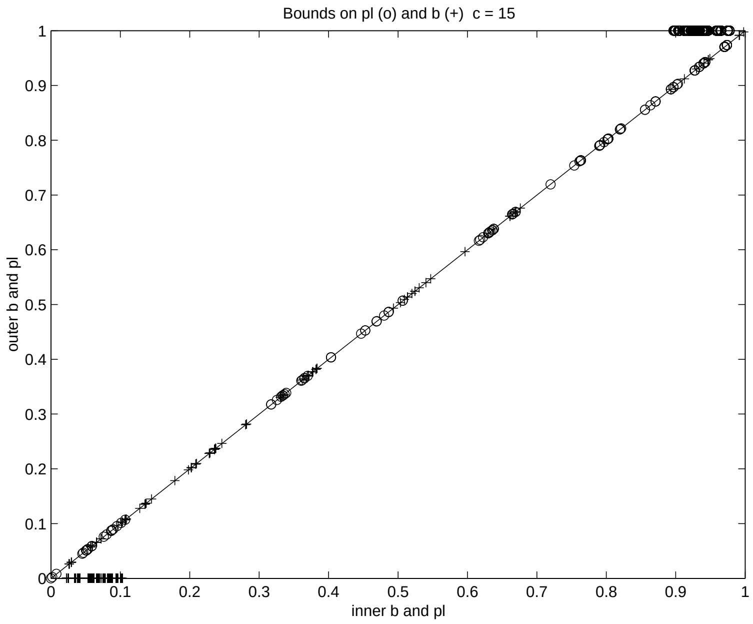


Figure 3: Inner vs. outer approximations of functions $b$ and pl (resulting from the conjunctive combination of 3 bba's) for $N = 1000$ randomly generated subsets of $\Omega$, with $|\Omega = 30|$, using a coarsening of size $c = |\Theta| = 15$.

shown by the symbols 'o' situated above the diagonal), and $\underline{b}^{\Omega}(A) \geq \overline{b}^{\Omega}(A)$ (as shown by the symbols '+' situated below the diagonal). The quality of the approximation is reflected by the proximity of the data points to the diagonal. As shown in Figure 2, the length of the plausibility and implicability intervals is often less 0.1 for $c = 10$, with some exceptions represented by the points located on the $y = 0$ and $y = 1$ lines, which correspond to significantly larger intervals. With $c = 15$ focal elements, the quality of the approximation is greatly improved (Figure 3).

In order to investigate the relationship between the size of the coarsening and the quality of the approximation, the following error measure was considered:

$$E = \frac{1}{2^n} \sum_{A \subseteq \Omega} \left( \overline{\mathrm{pl}}^{\Omega}(A) - \underline{\mathrm{pl}}^{\Omega}(A) \right) .$$

This quantity was estimated using $N = 1000$ randomly generated subsets $A_i, i = 1, \ldots, N$ of $\Omega$ by

$$\widehat{E} = \frac{1}{N} \sum_{i=1}^{N} \left( \overline{\mathrm{pl}}^{\Omega}(A_i) - \underline{\mathrm{pl}}^{\Omega}(A_i) \right) .$$

$E$ may be interpreted as an approximate upper bound of the mean approximation

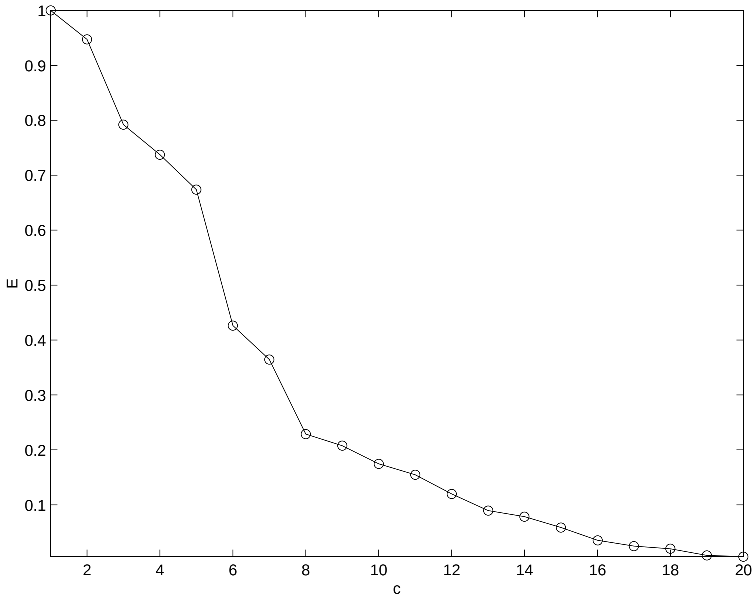


Figure 4: Average over 10 trials of the estimated approximation erreur $\widehat{E}$ (using 1000 randomly generated subsets), as a function of the coarsening size $c$.

error on the plausibilities, since we have

$$\overline{\mathrm{pl}}^{\Omega}(A) - \mathrm{pl}^{\Omega}(A) \leq \overline{\mathrm{pl}}^{\Omega}(A) - \underline{\mathrm{pl}}^{\Omega}(A)$$

$$\mathrm{pl}^{\Omega}(A) - \underline{\mathrm{pl}}^{\Omega}(A) \leq \overline{\mathrm{pl}}^{\Omega}(A) - \underline{\mathrm{pl}}^{\Omega}(A)$$

for all $A \subseteq \Omega$. The size $c$ of the coarsening was varied from 1 to 20, and the experiment was repeated 10 times with 10 different randomly generated triples of bba's. Figure 4 shows the average of $\widehat{E}$ over the 10 trials, as functions of $c$. As expected, the quality of the approximation increases with $c$, and it becomes almost perfect with $c = 20$. In general, the choice of $c$ should be guided by practical considerations, depending on the time and resources available for the computation.

**Remark 6** *This section was only intended to show typical results of our method applied to a "reasonable" approximation problem. Note that we have not attempted to perform any comparison with other methods, because there does not seem to be any standard way of artificially generating belief functions, nor is there any evidence of generic forms of belief functions being found in different applications. Consequently, it seems almost impossible to ensure that experimental results are not biaised in favor of* *a given approach, through the choice of a particular simulation model or a particular real-world application. This problem might be solved in the future if a large collection of realistic benchmark problems involving belief functions became available.*

## 5 Conclusion

A new method for computing strong inner and outer approximations of combined BF's has been defined. Unlike previous approaches, this method does not rely on the reduction of the number of focal elements, but on the construction of a coarsened frame in which combination can be performed efficiently using the FMT algorithm. Hence, this method makes the FMT algorithm usable for finding lower and upper bounds of combined implicability, belief and plausibility values, even for very large frames. When the number of focal elements is very large, such an approach may be expected to be more efficient than existing approximation methods, which rely exclusively on the mass-based combination algorithm.

If bba's are represented in matrix form, the method described can be seen as a hierarchical clustering procedure applied to the columns of a binary data matrix, with a particular dissimilarity measure. This contrasts with a method described in a previous paper [2], in which a similar algorithm was applied to the rows of the same matrix, each row corresponding to a focal element. Joint strategies aiming at reducing the number of focal elements and/or the size of the frame, depending on the problem at hand, could be considered as well, and are left for further study.


## Acknowledgements

The authors would like to express their thanks to the anonymous referees for their careful reading of the manuscript and for their valuable comments.


## References


[1] M. Bauer. Approximation algorithms and decision making in the Dempster-Shafer theory of evidence – an empirical study. *International Journal of Approximate*

*Reasoning*, 17:217–237, 1997.

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

[3] D. Dubois and H. Prade. Representation and combination of uncertainty with belief functions and possibility measures. *Computational Intelligence*, 4:244–264, 1988.

[4] D. Dubois and H. Prade. Consonant approximations of belief measures. *International Journal of Approximate Reasoning*, 4:419–449, 1990.

[5] D. Harmanec. Faithful approximations of belief functions. In K. B. Laskey and H. Prade, editors, *Uncertainty in Artificial Intelligence 15 (UAI99)*, Stockholm, Sweden, 1999.

[6] Lowrance J. D, T. D. Garvey, and T. M. Strat. A framework for evidential-reasoning systems. In T. Kehler et al., editor, *Proceedings of AAAI'86*, volume 2, pages 896–903, Philadelphia, August 1986. AAAI.

[7] A. K. Jain and R. C. Dubes. *Algorithms for clustering data.* Prentice-Hall, Englewood Cliffs, NJ., 1988.

[8] R. Kennes. Computational aspects of the Möbius transform of graphs. *IEEE Trans. SMC*, 22:201–223, 1992.

[9] R. Kennes and P. Smets. Computational aspects of the Möbius transformation. In P.P. Bonissone, M. Henrion, L. N. Kanal, and J. F. Lemmer, editors, *Uncertainty in Artificial Intelligence 6*, pages 401–416. Elsevier Science Publishers, 1991.

[10] G. J. Klir and M. J. Wierman. *Uncertainty-Based Information. Elements of Generalized Information Theory.* Springer-Verlag, New-York, 1998.

[11] P. Orponen. Dempster's rule of combination is # p-complete. *Artificial Intelligence*, 44:245–253, 1990.

[12] N. R. Pal, J. C. Bezdek, and R. Hemasinha. Uncertainty measures for evidential reasoning I: A review. *International Journal of Approximate Reasoning*, 7:165–183, 1992.

[13] N. R. Pal, J. C. Bezdek, and R. Hemasinha. Uncertainty measures for evidential reasoning II: New measure of total uncertainty. *International Journal of Approximate Reasoning*, 8:1–16, 1993.

[14] S. Petit-Renaud and T. Denœux. Handling different forms of uncertainty in regression analysis: a fuzzy belief structure approach. In A. Hunter and S. Pearsons, editors, *Symbolic and quantitative approaches to reasoning and uncertainty (ECSQARU'99)*, pages 340–351, London, June 1999. Springer Verlag.

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

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

[17] P. Smets. The Transferable Belief Model for quantified belief representation. In D. M. Gabbay and P. Smets, editors, *Handbook of Defeasible reasoning and uncertainty management systems*, volume 1, pages 267–301. Kluwer Academic Publishers, Dordrecht, 1998.

[18] P. Smets. Practical uses of belief functions. In K. B. Laskey and H. Prade, editors, *Uncertainty in Artificial Intelligence 15 (UAI99)*, pages 612–621, Stockholm, Sweden, 1999.

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

[20] N. Wilson. Algorithms for Dempster-Shafer theory. In D. M. Gabbay and P. Smets, editors, *Hanbook of defeasible reasoning and uncertainty management. Volume 5: Algorithms for Uncertainty and Defeasible Reasoning*, pages 421–475. Kluwer Academic Publishers, Boston, 2000.