New Basic belief assignment approximations based on optimization

Deqiang Han
Inst. of Integrated Automation
Xi'an Jiaotong University
Xi'an, China 710049
Email: deqhan@gmail.com

Jean Dezert
ONERA
The French Aerospace Lab
F-91761 Palaiseau, France
Email: jean.dezert@onera.fr

Chongzhao Han
Inst. of Integrated Automation
Xi'an Jiaotong University
Xi'an, China 710049
Email: czhan@mail.xjtu.edu.cn

***Abstract*—The theory of belief function, also called Dempster-Shafer evidence theory, has been proved to be a very useful representation scheme for expert and other knowledge based systems. However, the computational complexity of evidence combination will become large with the increasing of the frame of discernment's cardinality. To reduce the computational cost of evidence combination, the idea of basic belief assignment (bba) approximation was proposed, which can reduce the complexity of the given bba's. To realize a good bba approximation, the approximated bba should be similar (in some sense) to the original bba. In this paper, we use the distance of evidence together with the difference between the uncertainty degree of approximated bba and that of the original one to construct a comprehensive measure, which can represent the similarity between the approximated bba and the original one. By using such a comprehensive measure as the objective function and by designing some constraints, the bba approximation is converted to an optimization problem. Comparative experiments are provided to show the rationality of the construction of comprehensive similarity measure and that of the constraints designed.**

***Index Terms*—Evidence theory, belief function, bba, bba approximation, optimization.**

I. Introduction

The theory of belief function [1], which is also called Dempster-Shafer evidence theory (DST), provides an interesting and useful computational scheme for representing and integrating (or fusing) uncertain information. DST has been widely used in many applications, e.g., information fusion, pattern recognition and decision making [2]. However, high computational cost of evidence combination is a drawback which is often raised against DST [2]. It is well known that the computational cost of evidence combination increases exponentially with respect to cardinality of the frame of discernment (FOD) [3]- [5].

Many approaches have been proposed by the researchers to reduce the computational cost caused by evidence combination. Some researchers proposed efficient procedures for performing exact computations. For example, Kennes [6] proposed an optimal algorithm for Dempster's rule of combination. Barnett's work [7] and other works [8] are also the representatives. Moral and Salmeron [9] proposed the approach based on Monte-Carlo techniques. Wickramarathne also proposed bba approximation approaches based Monte-Carlo and statistical sampling [10]. Another important idea to reduce the computational cost caused by evidence combination is to approximate (or simplify) a bba to a simpler one. The papers of Voorbraak [4], Dubois and Prade [11] are seminal works in this type of approaches. Tessem proposed the famous $k-l-x$ approximation approach [3]. Grabisch proposed some approaches [12], which can build a bridge between belief functions and other types of uncertainty measures or functions, e.g., probabilities, possibilities and k-additive belief function (those belief functions whose cardinality of the focal elements are at most of k). Based on pignistic transformation in transferable belief model (TBM) [13], Burger and Cuzzolin proposed two types of k-additive belief functions [14]. Denœux uses hierarchical clustering strategy to implement the inner and outer approximation of belief functions [15]. In our recent research, a hierarchical proportional redistribution (HPR) approach is proposed to approximate bba [16].

In this paper, we focus on the bba approximation approach to reduce the computational cost of evidence combination. Although there have emerged many bba approximation approaches, we aim to design a new approach which can generate the optimal approximated bba using a reasonable criterion. A good approximated bba should have a structure (core) simpler than the original one, and at the same time it should be as similar as possible to the original bba. To obtain such a good approximated bba, we preset the maximum size of the focal element of the approximated bba and attempt to approximate bba by minimizing the dissimilarity between the original bba and the approximated one in this paper. Distance of evidence [17] is used together with the difference between the uncertainty degree of the approximated bba and that of the original one to measure the dissimilarity, which is used as the objective function of the optimization (minimization) problem. Different constraints for the optimization are also analyzed in this paper. Some examples are provided to show the rationality of the objective function and constraints used and to compare different bba approximation approaches. It can be experimentally shown that our proposed bba approximation approach is rational and efficient with respect to other approaches.

II. Basics of Belief Function Theory

Let $\Theta = \{\theta_1, ..., \theta_n\}$ be the frame of discernment (FOD), which is a set of exhaustive and mutually exclusive hypotheses. $\emptyset$ denotes the empty set. If $m : 2^\Theta \to [0,1]$ satisfies the requirements of $m(\emptyset)=0$ and $\sum_{A\subseteq\Theta} m(A)=1$, m is called the basic belief assignment (bba, or mass function) over the FOD Θ [1]. The belief function (Bel) and the plausibility function (Pl) are defined below [1], respectively:

$$Bel(A) = \sum\nolimits_{B\subseteq A} m(B) \quad (1)$$

$$Pl(A) = \sum\nolimits_{A\cap B\neq\emptyset} m(B) \quad (2)$$

In DST framework, Dempster's rule of combination was proposed for combining two distinct bodies of evidence (BOEs) characterized by bba's $m_1(.)$ and $m_2(.)$. Mathematically this rule is defined by:

$$m(A) = \begin{cases} 0, \ A=\emptyset \\ \frac{\sum\limits_{A_i\cap B_j=A} m_1(A_i)m_2(B_j)}{1-K}, \ A\neq\emptyset \end{cases} \quad (3)$$

where coefficient $K=\sum_{A_i\cap B_j=\emptyset} m_1(A_i)m_2(B_j)$ represents the conflict between the BOEs.

Dempster's rule of combination is both associative and commutative and can be extended for combining $n>2$ distinct sources of evidence as well. Aside debates on its validity, Dempster's rule of combination requires significantly large computational cost with the increasing of the FOD's cardinality. Given a FOD with cardinaltiy of n, bba $m(.)$ can have up to 2^n-1 focal elements (empty-set $\emptyset$ here is excluded), Furthermore, the combination of two bba's requires the computation of up to $2^{n+1}-2$ intersections.

To work with large FOD and make DST tractable, several approaches have been proposed either by proposing efficient algorithms [6]- [8] for evidence combination, by using Monte-Carlo techniques, or by approximating bba[9], [10]. We prefer to use the bba approximation approach [11] - [14] to reduce the computational cost needed in the combination operation because the approximation approach reduces the computational cost and also allow to deal with smaller-size focal elements, which is more intuitive for human to catch the meaning [18].

III. Recent bba approximation approaches

For the purpose of comparisons, some recent bba approximation approaches are recalled in this section.

1) k-additive belief function approximation: Given a bba $m : 2^\Theta \to [0,1]$, the k-additive belief function [12] induced by the mass assignment is defined in Eq.(5). Suppose that $B\subseteq\Theta$,

$$\begin{cases} m_k(B) = m(B) + \sum_{\substack{A\supset B\\ A\subseteq\Theta\\ |A|>k}} \frac{m(A)\cdot|B|}{\mathcal{N}(|A|,k)}, & \forall\,|B|\leq k \\ m_k(B)=0, & \forall\,|B|>k \end{cases} \quad (4)$$

where

$$\mathcal{N}(|A|,k) = \sum_{j=1}^{k} \binom{|A|}{j}\cdot j = \sum_{j=1}^{k} \frac{|A|!}{(j-1)!(|A|-j)!} \quad (5)$$

is average cardinality of the subsets of A of size at most k.

It can be seen that for k-additive belief approximation, the maximum cardinality of available focal elements is no greater than k.

2) Hierarchical Proportional Redistribution approximation: In our previous research work [16], we have proposed a new bba approximation approach called hierarchical proportional redistribution (HPR), which provides a new way to reduce step-by-step the mass committed to uncertainties. Ultimately an approximate measure of subjective probability can be obtained if needed, i.e. a so-called Bayesian bba. The HPR procedure can be stopped at any step in the process and thus it allows to reduce the number of focal elements of a given bba in a simple manner to diminish the size of the core of a bba. Thus we can reduce the complexity (if needed) when applying also some complex rules of combinations. By using HPR, we can obtain approximate bba's at any different non-specificity level that we want. Let's first introduce two new notations for convenience and for concision:

1) Any element of cardinality $1\leq k\leq n$ of the power set 2^Θ will be denoted, by convention, by the generic notation $X(k)$. For example, if $\Theta=\{A,B,C\}$, then $X(2)$ can denote the following partial uncertainties $A\cup B$, $A\cup C$ or $B\cup C$, and $X(3)$ denotes the total uncertainty $A\cup B\cup C$.
2) The proportional redistribution factor (ratio) of width n involving elements X and Y of the powerset is defined as (for $X\neq\emptyset$ and $Y\neq\emptyset$)

$$R_s(Y,X) \triangleq \frac{m(Y)+\epsilon\cdot|X|}{\sum_{\substack{Y\subset X\\ |X|-|Y|=s}} m(Y)+\epsilon\cdot|X|} \quad (6)$$

where ϵ is a small positive number introduced here to deal with particular cases where $\sum_{\substack{Y\subset X\\ |X|-|Y|=s}} m(Y)=0$.

By convention, we will denote $R(Y,X)\triangleq R_1(Y,X)$ when we use the proportional redistribution factors of width $s=1$.

The HPR is obtained by a step by step (recursive) proportional redistribution of the mass $m(X(k))$ of a given uncertainty $X(k)$ (partial or total) of cardinality $2\leq k\leq n$ to all the least specific elements of cardinality $k-1$, i.e., to all possible $X(k-1)$, until $k=2$ is reached. The proportional redistribution is done from the masses of belief committed to $X(k-1)$ as done classically in DSmP transformation. The "hierarchical" masses $m_h(.)$ are recursively (backward) computed as follows. Here $m_{h(k)}$ represents the approximate bba obtained at the step $n-k$ of HPR, i.e., it has the maximum focal element cardinality of k.

$$\begin{aligned} m_{h(n-1)}(X(n-1)) &= m(X(n-1)) + \\ &\sum_{\substack{X(n)\supset X(n-1)\\ X(n),X(n-1)\in 2^\Theta}} [m(X(n))\cdot R(X(n-1),X(n))]; \\ m_{h(n-1)}(A) &= m(A), \forall|A|<n-1 \end{aligned} \quad (7)$$

$m_{h(n-1)}(\cdot)$ is the bba obtained at the first step of HPR ($n-(n-1)=1$), the maximum focal element cardinality of $m_{h(n-1)}$ is $n-1$.

$$m_{h(n-2)}(X(n-2)) = m(X(n-2)) + \sum_{\substack{X(n-1)\supset X(n-2)\\ X(n-2),X(n-1)\in 2^\Theta}} [m_{h(n-1)}(X(n-1))\cdot R(X(n-2),X(n-1))]$$

$$m_{h(n-2)}(A) = m_{h(n-1)}(A), \forall |A| < n-2 \tag{8}$$

$m_{h(n-2)}(\cdot)$ is the bba obtained at the second step of HPR ($n-(n-2)=2$), the maximum focal element cardinality of $m_{h(n-2)}$ is $n-2$.

This hierarchical proportional redistribution process can be applied similarly (if one wants) to compute $m_{h(n-3)}(.)$, $m_{h(n-4)}(.)$, ..., $m_{h(2)}(\cdot)$, $m_{h(1)}(\cdot)$ with

$$m_{h(2)}(X(2)) = m(X(2)) + \sum_{\substack{X(3)\supset X(2)\\ X(3),X(2)\in 2^\Theta}} [m_{h(3)}(X(3))\cdot R(X(2),X(3))]$$

$$m_{h(2)}(A) = m_{h(3)}(A), \forall |A| < n-2 \tag{9}$$

$m_{h(2)}(\cdot)$ is the bba obtained at the first step of HPR ($n-2$), the maximum focal element cardinality of $m_{h(2)}$ is 2.

Mathematically, for any $X(1)\in\Theta$, i.e. any $\theta_i\in\Theta$ a Bayesian belief function can be obtained by HPR approach in deriving all possible steps of proportional redistributions of partial ignorances in order to obtain

$$m_{h(1)}(X(1)) = m(X(1)) + \sum_{\substack{X(2)\supset X(1)\\ X(1),X(2)\in 2^\Theta}} [m_{h(2)}(X(2))\cdot R(X(1),X(2))] \tag{10}$$

In fact, $m_{h(1)}(\cdot)$ is a probability transformation, called here the Hierarchical DSmP (HDSmP). Since $X(n)$ is unique and corresponds only to the full ignorance $\theta_1\cup\theta_2\cup\ldots\cup\theta_n$, the expression of $m_h(X(n-1))$ in Eq.(7) just simplifies as

$$m_{h(n-1)}(X(n-1)) = m_h(X(n-1)) + m(X(n))\cdot R(X(n-1),X(n)) \tag{11}$$

For the full proportional redistribution of the masses of uncertainties to the elements least specific involved in these uncertainties, no mass is lost during the step-by-step hierarchical process and thus at any step of HPR, the sum of masses is kept to one. Illustrative examples of HPR can be found in [16].

IV. BBA APPROXIMATIONS BASED ON OPTIMIZATION

A "good" approximated bba should be simpler[1] than the original one and at the same time should be as similar as possible to the original bba. To obtain such a "good" approximated bba, in this paper we predefine the maximum focal element size of the approximated bba as k to ensure the "simplicity" and we attempt to approximate bba by minimizing the dissimilarity between the original one and the approximated one to ensure the "similarity". That is to say, the problem of bba approximation is converted to an optimization problem. For the optimization problem, we need to define the objective function and some constraints which will be detailed in the next subsections.

A. Design of objective function

Our purpose is to minimize the dissimilarity between the original bba and the approximated one. But how to measure the dissimilarity between bba's? The distance of evidence seems to be a good primary choice and several types of distance of evidence have been proposed in the literature (see Jousselme et al's latest survey in [19] for details). Among all the proposals for distances of evidence, we have chosen Jousselme's distance based on the form of Euclidean metric because it is a strict distance metric [19] and takes into account the specificity of focal[2] elements of the bba. This distance is defined by

$$d_J(m_1,m_2) = \sqrt{(m_1-m_2)^T \mathbf{Jac}\ (m_1-m_2)} \tag{12}$$

Jac is the Jaccard's weighting matrix whose elements are given by

$$\mathbf{Jac}(A,B) = \frac{|A\cap B|}{|A\cup B|} \tag{13}$$

where A and B represent the focal elements of $m_1(.)$ and $m_2(.)$, respectively.

Aside Jousselme's distance, we propose to use also the difference between the uncertainty degree of two different bba's to describe their dissimilarity. Several measures of uncertainty have been proposed in the literature and we just recall the most important ones below:

1) Aggregated Uncertainty (AU): Let $m(.)$ be a bba on the FOD Θ. The AU associated with $m(.)$ is defined by [20]:

$$\text{AU}(m) = \max_{\mathcal{P}_m}[-\sum_{\theta\in\Theta} p_\theta \log_2 p_\theta] \tag{14}$$

where the maximum is taken over all probability distributions that are consistent with the given bba. $\mathcal{P}_m$ consists of all probability distributions $\langle p_\theta|\theta\in\Theta\rangle$ satisfying the constraints:

$$\begin{cases} p_\theta\in[0,1], \forall\theta\in\Theta \\ \sum_{\theta\in\Theta} p_\theta = 1 \\ Bel(A) \le \sum_{\theta\in A} p_\theta \le 1 - Bel(\bar{A}),\ \forall A\subseteq\Theta \end{cases} \tag{15}$$

[1]The core of approximated bba must be smaller than the core of the original bba.

[2]X is a focal element of a bba $m(.)$ if $m(X)>0$.

As illustrated in Eq. (14) and Eq. (15), in the definition of AU, the calculation of AU is an optimization problem and bba's (or belief functions) are used to establish the constraints of the optimization problem. It is also called the "upper entropy". AU is a strict uncertainty measure satisfying the five requirements for uncertainty measure [20]. However, AU has a high computing complexity and a high insensitivity to the changes of evidence [21] which makes this measure not very attractive.

2) Ambiguity measure (AM): Let $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ be a FOD with n elements. Let m be a bba defined on Θ. The Ambiguity Measure (AM) [21] is defined by

$$\text{AM(m)} = -\sum\nolimits_{\theta\in\Theta} \text{BetP}_m(\theta)\log_2(\text{BetP}_m(\theta)) \tag{16}$$

where $\text{BetP}_m(\{\theta\}) = \sum_{\theta\in B, B\subseteq\Theta} m(B)/|B|$ is the pignistic probability distribution [13] computed from the bba $m(.)$. In fact AM does not satisfy the subadditivity as pointed out by Klir in [22]. Also AM is logically non-monotonic under some circumstances as proved by Abellan [20]. AU and AM are both established based on Shannon entropy over some probability transformation resulting from a given bba.

3) Non-specificity: This measure introduced in [23] and denoted $U(m)$ is defined directly from $m(.)$ by

$$U(m) = \sum_{A\subseteq\Theta} m(A)\log_2|A| \tag{17}$$

It is worth noting that this measure describes the non-specificity aspect (when two or more alternatives left unspecified) of the uncertain degree incorporated in a body of evidence but cannot well discriminate two distinct bba's in some cases. For example, let's consider the two bba's $m_1(.)$ and $m_2(.)$ given by

$$\begin{cases} m_1(\theta_1\cup\theta_2) = 0.1, m_1(\theta_2\cup\theta_3) = 0.8, m_1(\theta_1\cup\theta_3) = 0.1 \\ m_2(\theta_1\cup\theta_2) = 1/3, m_2(\theta_2\cup\theta_3) = 1/3, m_2(\theta_1\cup\theta_3) = 1/3 \end{cases}$$

It is clear that their corresponding non-specificity values are the same: $U(m_1) = U(m_2) = 1$, although $m_1(.)$ and $m_2(.)$ are obviously different.

As already mentioned, AU and AM have their drawbacks. Because AU and AM are established based on some bba approximations (in fact, probability transformations), so there exist the loss of information and that is why we don't recommend them as efficient measures of uncertainty degree of a body of evidence. Although the non-specificity measure $U(.)$ represents only one aspect of the uncertainty, it is defined directly based on bba and it is easy to calculate. For these reasons, we prefer to use the non-specificity jointly with Jousselme's distance to construct the objective function in our optimization problem.

The construction of our objective function to minimize is based both on Jousselme's distance $d_J(.,.)$ and the non-specificity measure $U(.)$ as follows:

1) we use distance of evidence in Eq. (12) to construct the objective function

$$\text{obj}_1(m) = d_J(m, m_{org})$$

2) we use the non-specificity to construct the objective function

$$\text{obj}_2(m) = \frac{U(m_{org}) - U(m)}{U(m_{org})}$$

where $m_{org}(.)$ denotes the original bba to be approximated by the bba $m(.)$.

In this paper, we propose to use these measures of distance of evidence and non-specificity to construct the global objective function to minimize. How to combine them into one comprehensive global objective function? Since one does not have a priori clear answer to this problem, we have examined the simplest two methods for doing this.

1) Additive global objective: It is defined by

$$\text{obj}(m) = \alpha\cdot\text{obj}_1(m) + \beta\cdot\text{obj}_2(m)$$

More precisely, by

$$\text{obj}(m) = \alpha\cdot d_J(m_{org}, m) + \beta\cdot\frac{U(m_{org}) - U(m)}{U(m_{org})} \tag{18}$$

where $\alpha, \beta \in [0, 1]$ represent the importance or degree of preference of obj_1 and obj_2, respectively.

2) Multiplicative global objective: It is defined by

$$\text{obj}(m) = (\text{obj}_1(m))^\alpha\cdot(\text{obj}_2(m))^\beta$$

More precisely, by

$$\text{obj}(m) = (d_J(m_{org})^\alpha\cdot(\frac{U(m_{org}) - U(m)}{U(m_{org})})^\beta \tag{19}$$

B. Design of constraints

1) Weak constraints: The approximated bba $m(.)$ should at first satisfy the definition of bba. Suppose the maximum cardinality of the approximated bba $m(.)$ is k, then we must have

$$\begin{cases} 0 \leq m(A) \leq 1, \forall A\subseteq\Theta \\ \sum_{\substack{B\subseteq\Theta \\ |B|\leq k}} m(B) = 1 \end{cases} \tag{20}$$

Because the purpose of the bba approximation is to reduce the complexity of a given bba, the uncertainty degree of the approximated bba should not be higher than the original one. This principle introduces naturally the constraint $U(m) \leq U(m_{org})$. Therefore the constraints that the approximated bba $m(.)$ must satisfy are

$$\begin{cases} 0 \leq m(A) \leq 1, \forall A\subseteq\Theta \\ \sum_{\substack{B\subseteq\Theta \\ |B|\leq k}} m(B) = 1 \\ U(m) < U(m_{org}) \end{cases} \tag{21}$$

These constraints above are in fact too weak as shown in Example 1.

Example 1: Suppose that the FOD is $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_4\}$ and the given bba is

$m_{org}(\theta_1) = 0.8, m_{org}(\theta_2) = 0.04, m_{org}(\theta_3) = 0.04;$

$m_{org}(\theta_4) = 0.04, m_{org}(\theta_2 \cup \theta_3) = 0.04,$

$m_{org}(\{\theta_2 \cup \theta_3 \cup \theta_4\}) = 0.04$

If we use the constraints in Eq. (21) and the minimization of the global multiplicative objective, we obtain the following approximated bba

$m(\theta_1) = 0.7850, m(\theta_2) = 0.0272, m(\theta_3) = 0.0272;$

$m(\theta_4) = 0.0573, m(\theta_2 \cup \theta_3) = 0.1034;$

One sees that we have also diluted (diminished) the precise information focused on θ_1 only (0.8), focused on θ_2 only (0.04), focused on θ_3 only (0.04) and focused on θ_4 only (0.04) to other partial uncertainty.

If we use the constraints in Eq. (21) and according to the minimization of the global additive objective ($\alpha = \beta = 1$), we obtain the approximated bba as

$m(\theta_1) = 0.8000, m(\theta_2) = 0.0216, m(\theta_3) = 0.0216;$

$m(\theta_4) = 0.0533, m(\theta_2 \cup \theta_3) = 0.1034;$

Based on the minimization of the global additive objective under constraints, one has also diluted the precise information focused on θ_2 only (0.04) and focused on θ_3 only (0.04) to other partial uncertainty. In both cases (with additive or multiplicative global objectives), we don't see any serious and legitimate reasons for justifying such behavior in practice. There exists another extreme case as shown in Example 2.

Example 2: Suppose that the FOD is $\Theta = \{\theta_1, \theta_2, \theta_3\}$ and the given bba is

$m_{org}(\theta_1) = 0.1, m_{org}(\theta_2) = 0.1, m_{org}(\theta_3) = 0.0;$

$m_{org}(\theta_1 \cup \theta_2) = 0.3, m_{org}(\theta_2 \cup \theta_3) = 0.1$

$m_{org}(\theta_1 \cup \theta_2 \cup \theta_3) = 0.4$

If we minimize the global additive (or multiplicative) objective under the constraints in Eq. (21) with $\alpha = \beta = 1$, we obtain the following approximated bba

$m(\theta_1) = 0.0, m(\theta_2) = 0.0, m(\theta_3) = 0.0;$

$m(\theta_1 \cup \theta_2) = 0.6375, m(\theta_2 \cup \theta_3) = 0.3625$

and one sees that the mass assignments for singletons are completely diluted. There is no legitimate reason for doing that. This phenomenon is due to the weakness of the constraints used. Therefore to circumvent the problem, we need to consider stronger constraints.

2) Strong constraints: To be assured that there is no dilution phenomenon as illustrated in Examples 1 or 2, we add the following additional constraint

$$m(A) \geq m_{org}(A), \forall |A| \leq k \tag{22}$$

where k is the maximum allowed focal element size of the approximated bba. We thus prevent the dilution by doing this. Then the enforced set of constraints is

$$\begin{cases} m(A) \geq m_{org}(A), \forall |A| \leq k \\ 0 \leq m(A) \leq 1, \forall A \subseteq \Theta \\ \sum_{A \subseteq \Theta} m(A) = 1, \forall |A| \leq k \end{cases} \tag{23}$$

If we use the strong constraints (23) in Example 1 for bba approximation, we obtain the following bba[3]

$m(\theta_1) = 0.8, m(\theta_2) = 0.04, m(\theta_3) = 0.04;$

$m(\theta_3) = 0.04, m(\theta_4) = 0.04, m(\theta_2 \cup \theta_3) = 0.08.$

If we use the strong constraints (23) in Example 2 for bba approximation, we obtain the following bba[4]

$m(\theta_1) = 0.1, m(\theta_2) = 0.1, m(\theta_3) = 0.0;$

$m(\theta_1 \cup \theta_2) = 0.5, m(\theta_2 \cup \theta_3) = 0.3$

As we can see, there is no phenomena of dilution. The strong constraints are more rational than the weak constraints.

C. *In summary*

Based on the strong constraints and by using either the global additive or the global multiplicative objectives, we can establish two optimization problems for bba approximations which can be summarized as follows:

Optimization_additive

$$\begin{aligned} &\min_m \left(\text{obj}(m) = \alpha \cdot d_J(m_{org}, m) + \beta \cdot \tfrac{U(m_{org}) - U(m)}{U(m_{org})} \right) \\ &s.t. \begin{cases} m(A) \geq m_{org}(A), \forall |A| < k \\ 0 \leq m(A) \leq 1, \forall A \subseteq \Theta \\ \sum_{\substack{B \subseteq \Theta \\ |B| \leq k}} m(B) = 1 \end{cases} \end{aligned} \tag{24}$$

and

Optimization_multiplicative

$$\begin{aligned} &\min_m \left(\text{obj}(m) = (d_J(m_{org}, m))^{\alpha} \cdot (\tfrac{U(m_{org}) - U(m)}{U(m_{org})})^{\beta} \right) \\ &s.t. \begin{cases} m(A) \geq m_{org}(A), \forall |A| \leq k \\ 0 \leq m(A) \leq 1, \forall A \subseteq \Theta \\ \sum_{\substack{B \subseteq \Theta \\ |B| \leq k}} m(B) = 1 \end{cases} \end{aligned} \tag{25}$$

V. EXPERIMENTS

In this section, we randomly generate bba's and compare different bba approximation approaches including the new proposed bba approximation approaches, k-additive and HPR approaches. The bba's are generated according to Algorithm 1 below [24]:

Algorithm 1. Random generation of bba

Input: Θ : Frame of discernment;
N_{max}: Maximum number of focal elements
Output: Bel: Belief function (in the form of a bba, m)
Generate $\mathcal{P}(\Theta)$, which is the power set of Θ;
Generate a random permutation of $\mathcal{P}(\Theta) \rightarrow \mathcal{R}(\Theta)$;
Generate an integer between 1 and $N_{max} \rightarrow l$;
FOReach First l elements of $\mathcal{R}(\Theta)$ do
Generate a value within $[0, 1] \rightarrow m_i$, $i = 1, ..., l$;
$m(A_i) = m_i$;
END
Normalize the vector $m = [m_1, ..., m_l] \rightarrow m'$;

[3] The same result is obtained with multiplicative or with additive global objective with $\alpha = \beta = 1$.

[4] idem as footnote 3.

In our tests, we have set the cardinality of the FOD to 5 and fixed the number of focal elements to $l = N_{max} = 2^5 - 1 = 31$. We randomly generate bba's for $L = 30$ times. At each time, different bba approximations are executed with remaining maximum focal element's size of 4, 3, and 2, respectively. Non-specificity and distance of evidence are used to evaluate their corresponding approximated bba's, which are shown in Fig. 1 and Fig. 2. The parameter of optimization problem (additive way and multiplicative way) are $\alpha = \beta = 1$

As we can see in Fig. 1 and Fig. 2, the optimization approach based on the global additive objective and the optimization approach based on the global multiplicative objective always have lowest distance values and have the non-specificity values being closest to the those of the original bba's. This means that our proposed optimization approach for bba approximation is rational and it is better than k-additive and HPR. However, according to the non-specificity criterion, HPR performs as good as our bba approximation approaches based on optimization of our objective functions.

It should be noted that the experimental results of the optimization approach based on global additive objective and based on global multiplicative objective are very close, so we plot two figures to respectively illustrate their comparisons with other approaches.

Furthermore, we have used also the normalized mean squared error (NMSE) criterion to evaluate their performances. NMSE is defined by

$$NMSE(error) = \frac{1}{L}\sum_{i=1}^{L}\frac{(\vec{e}(i))^2}{\text{variance}(\vec{e})} \quad (26)$$

where L is the number of tests, and the error vectors are defined for $i = 1, 2, \dots, L$ by

$$\vec{e(i)} = \text{obj}(m_{org}) - \text{obj}(m(i))$$

with $\text{obj}(\cdot)$ given by Eq. (18).

Our simulations results of NMSE obtained by the different approaches are list in Table I below.

Table I NMSE Comparisons among different approaches

Approaches	$k = 4$	$k = 3$	$k = 2$
Opt_add	13.5052	17.9532	43.2814
Opt_multi	13.5123	17.9533	43.2816
HPR	13.5715	18.1212	43.3821
k-additive	14.1915	20.4053	54.9697

As we can see in Table I, the optimization approach based on global additive objective always provides the lowest NMSE values, which indicates that the optimization approach based on global additive objective performs the best.

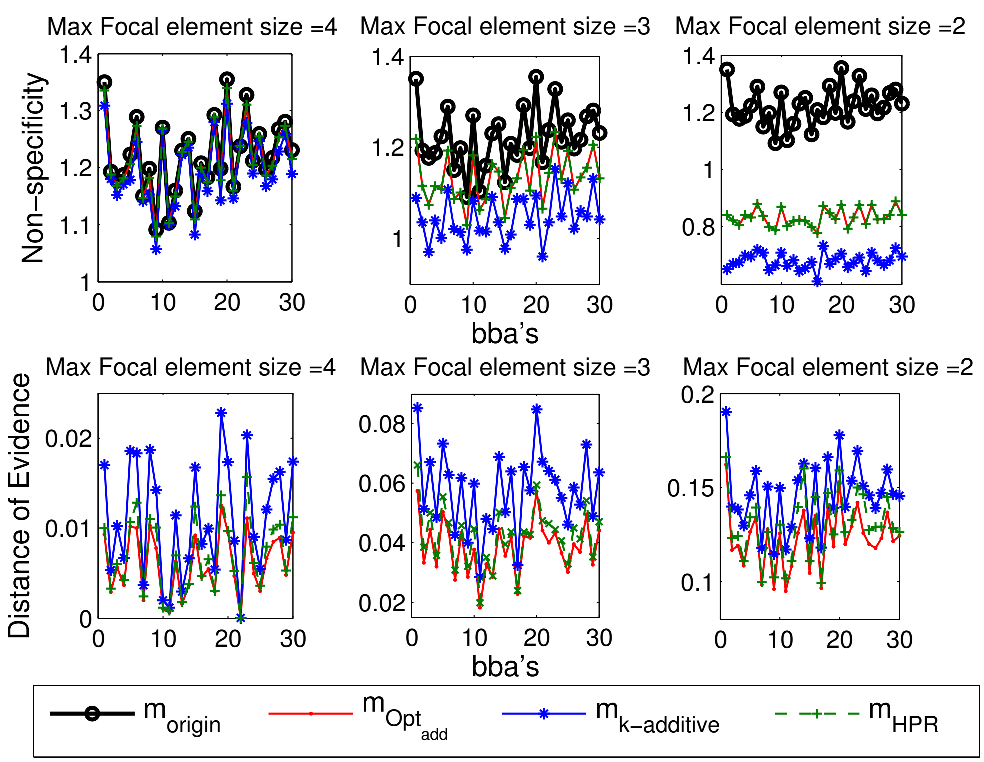

Fig. 1. Comparisons of Opt_additive with HPR and k-additive.

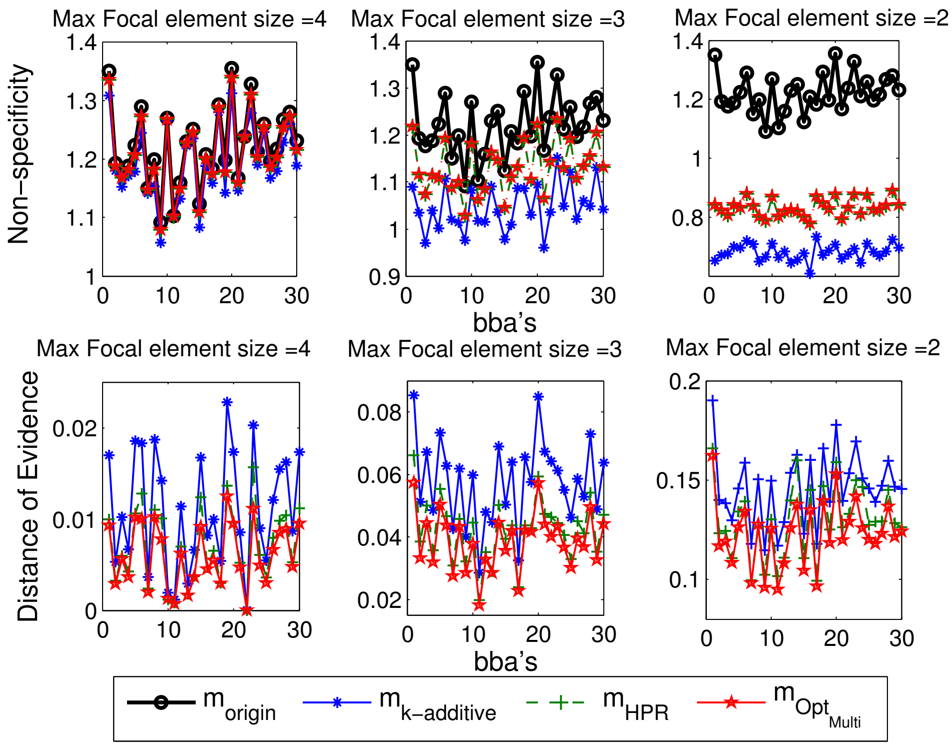

Fig. 2. Comparisons of Opt_multiplicative with HPR and k-additive.

VI. Conclusion

In this paper, new bba approximation approaches have been proposed based on optimization. The construction of objective functions and the constraints were also discussed and analyzed. Experimental results show the rationality of our proposed approaches. It should be noted that although our new approaches performs well, they are based on optimization, which will cause more computational complexity. That is to say the good performance is at the price of computational complexity. Our previously proposed HPR has comparable performances with respect to the performances of our new optimization approaches, but HPR does not has so high computational complexity as bba approximations based on optimization. So in some time-sensitive applications, HPR is a better choice. Furthermore, in our objective function construction, there still exists the problem of selection of the weighting parameters α and β. Their effects on the results of the bba approximations will be analyzed in details in future research works.

Acknowledgment

This work is supported by National Natural Science Foundation of China (Grant No.61104214, No.67114022, No. 61074176, No. 60904099), Fundamental Research Funds for the Central Universities, China Postdoctoral Science Foundation (No.20100481337, No.201104670), Research Fund of Shaanxi Key Laboratory of Electronic Information System Integration (No.201101Y17), Chongqing Natural Science Foundation, Grant No. CSCT, 2010BA2003 and Aviation Science Foundation (No. 2011ZC53041).

References

[1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University, Princeton, 1976.

[2] P. Smets, "Practical uses of belief functions", in K. B. Lskey and H. Prade, Editors, Uncertianty in Artificial Intelligence 15 (UAI 99), pages 612–621, Stockholm, Sweden, 1999.

[3] B. Tessem, "Approximations for efficient computation in the theory of evidence", Artificial Intelligence, Vol. 61, no. 2, pp. 315–329, June 1993.

[4] F. Voorbraak, "A computationally efficient approximation of Dempster-Shafer theory", International Journal of Man-Machine Studies, Vol. 30, no. 5, pp. 525–536, 1989.

[5] F. Smarandache, J. Dezert (Editors), *Applications and Advances of DSmT for Information Fusion*, Vol 3, American Research Press, 2009. http://www.gallup.unm.edu/˜smarandache/DSmT-book3.pdf.

[6] R. Kennes, "Computational aspects of the Möbius transform of graphs", IEEE Transactions on SMC, Vol. 22, no .2, pp. 201–223, 1992.

[7] J.A. Barnett, "Computational methods for a mathematical theory of evidence", in Proceedings of IJCAI-81, pp. 868–875, Vancouver, 1981.

[8] G. Shafer, R. Logan, "Implementing Dempster's rule for hierarchical evidence", Artificial Intelligence, Vol. 33, no .3, pp. 271–298, 1987.

[9] S. Moral, A. Salmeron, "A Monte Carlo algorithm for combining Dempster-Shafer belief based on approximate pre-computation", in A. Hunter and S. Pearsons, Editors, Symbolic and quantitative approaches to reasoning and uncertainty (ECSQARU'99), pp. 305–315, 1999.

[10] T. L. Wickramarathne, K. Premaratne, M. N. Murthi, "Monte-Carlo approximations for Dempster-Shafer belief theoretic algorithms", in Proc. 14th International Conference on Information Fusion, Chicago, USA, July 5-8, 2011, pp. 461–468.

[11] D. Dubois,H. Prade, "An alternative approach to the handling of subnormal possibility distributions", Fuzzy Sets and Systems, Vol. 24, no. 1, pp. 123–126, 1987.

[12] M. Grabisch, "Upper approximation of non-additive measures by k-additive measures - the case of belief functions", in Proc. of the 1st Int. Symposium on imprecise probabilities and their applications, Ghent, Belgium, June 1999.

[13] P. Smets, R. Kennes, "The transferable belief model", Artificial Intelligence, Vol 66, No. 2, pp. 191-234, 1994.

[14] T. Burger, F. Cuzzolin, "The barycenters of the k-additive dominating belief functions and the pignistic k-additive belief functions", Proc. of International Workshop on Belief Funcitons (BELIEF'10), Brest, France, April 1st - April 2nd.

[15] T. Denœux, "Inner and outer approximation of belief structures using a hierarchical clustering approach", International Journal of Uncertainty, Fuzziness, and Knowledge-based Systems, Vol. 9, no. 4, pp. 437–460, 2001.

[16] J. Dezert, D.Q. Han, Z.G. Liu, J.-M. Tacnet, "Hierarchical proportional redistribution for bba approximation", in Proceedings of the 2nd International Conference on Belief Functions, Compiègne, France, May 9-11, 2012.

[17] A.-L. Jousselme, D. Grenier, E. Bosse, "A new distance between two bodies of evidence", Information Fusion, Vol. 2, no. 2, pp. 91–101, June 2001.

[18] T. Burger, "Defining new approximations of belief functions by means of Dempster's combination", in Proc. of the 1st International Workshop on the Theories of Belief Functions (BELIEF 2010), Brest, France, April 1st - April 2nd, 2010.

[19] A.-L. Jousselme, P. Maupin, "Distances in evidence theory: Comprehensive survey and generalizations". International Journal of Approximate Reasoning, Vol. 53, no. 2, pp. 118–145, 2011

[20] J. Abellan and A. Masegosa. "Requirements For Total Uncertainty Measures In Dempster-shafer theory Of Evidence", International Journal of General Systems, Vol. 37, no. 6, pp. 733–747, 2008.

[21] A.-L. Jousselme, C. S. Liu, D. Grenier, E. Bosse, "Measuring ambiguity in the evidence theory", *IEEE Transactions on Systems, Man and Cybernetics, Part A*, vol. 36, no. 5, pp. 890–903, 2006.

[22] G. J. Klir and H. W. Lewis. Remarks on "Measuring Ambiguity in the Evidence Theory", IEEE Transactions on Systems, Man and Cybernetics, Part A, Vol. 38, no.4, pp. 995–999, 2008.

[23] D. Dubois, H. Prade, "A note on measures of specificity for fuzzy sets", International Journal of General Systems, Vol. 10, no. 4, pp. 279–283, 1985.

[24] A.-L. Jousselme, P. Maupin, "On some properties of distances in evidence theory," in *Workshop on Theory of Belief Functions*, Brest, France, Mar 31st - Apr 2nd, 2010, pp. 1-6.