Hierarchical DSmP Transformation for Decision-Making under Uncertainty

Jean Dezert

Deqiang Han

Zhun-ga Liu

Jean-Marc Tacnet

Abstract—**Dempster-Shafer evidence theory is widely used for approximate reasoning under uncertainty; however, the decision-making is more intuitive and easy to justify when made in the probabilistic context. Thus the transformation to approximate a belief function into a probability measure is crucial and important for decision-making based on evidence theory framework. In this paper we present a new transformation of any general basic belief assignment (bba) into a Bayesian belief assignment (or subjective probability measure) based on new proportional and hierarchical principle of uncertainty reduction. Some examples are provided to show the rationality and efficiency of our proposed probability transformation approach.**

Keywords: Belief functions, probabilistic transformation, DSmP, uncertainty, decision-making.

I. INTRODUCTION

Dempster-Shafer evidence theory (DST) [1] proposes a mathematical framework for approximate reasoning under uncertainty thanks to belief functions. Thus it is widely used in many fields of information fusion. As any theory, DST is not exempt of drawbacks and limitations, like its inconsistency with the probability calculus, its complexity and the miss of a clear decision-making process. Aside these weaknesses, the use of belief functions remains flexible and appealing for modeling and dealing with uncertain and imprecise information. That is why several modified models and rules of combination of belief functions were proposed to resolve some of the drawbacks of the original DST. Among the advances in belief function theories, one can underline the transferable belief model (TBM) [2] proposed by Smets, and more recently the DSmT [3] proposed by Dezert and Smarandache.

The ultimate goal of approximate reasoning under uncertainty is usually the decision-making. Although the decision-making can be done based on evidence expressed by a belief function [4], the decision-making is better established in a probabilistic context: decisions can be evaluated by assessing their ability to provide a winning strategy on the long run in a game theory context, or by maximizing return in a utility theory framework. Thus to take a decision, it is usually preferred to transform (approximate) a belief function into a probability measure. So the quality of such probability transformation is crucial for the decision-making in the evidence theory. The research on probability transformation has attracted more attention in recent years.

The classical probability transformation in evidence theory is the pignistic probability transformation (PPT) [2] in TBM. TBM has two levels: the credal level, and the pignistic level. Beliefs are entertained, combined and updated at the credal level while the decision making is done at the pignistic level. PPT maps the beliefs defined on subsets to the probability defined on singletons. In PPT, belief assignments for a compound focal element are equally assigned to the singletons included. In fact, PPT is designed according to the principle of minimal commitment, which is somehow related with uncertainty maximization.

Other researchers also proposed some modified probability transformation approaches [5]–[13] to assign the belief assignments of compound focal elements to the singletons according to some ratio constructed based on some available information. The representative transformations include Sudano's probability transformations [8] and Cuzzolin's intersection probability transformation [13], etc. In the framework of DSmT, another probability transformation approach was proposed, which is called DSmP [9]. DSmP takes into account both the values of the masses and the cardinality of focal elements in the proportional redistribution process. DSmP can also be used in both DSmT and DST. For a probability transformation, it is always evaluated by using probabilistic information content (PIC) [5] (PIC being the dual form of Shannon entropy), although it is not enough or comprehensive [14]. A probability transformation providing a high probabilistic information content (PIC) is preferred in fact for decision-making since naturally it is always easier to take a decision when the uncertainty is reduced.

In this paper we propose a new probability transformation, which can output a probability with high but not exaggerated PIC. The new approach, called HDSmP (standing for Hierarchical DSmP) is implemented hierarchically and it fully utilize the information provided by a given belief function. Succinctly, for a frame of discernment (FOD) with size n, for $k = n$ down to $k = 2$, the following step is repeated: the belief assignment of a focal element with size k is proportionally redistributed to the focal elements with size $k - 1$. The proportion is defined by the ratio among mass assignments of focal elements with size $k - 1$. A parameter ϵ is introduced in the formulas to avoid division by zero and warranty numerical

robustness of the result. HDSmP corresponds to the last step of the hierarchical proportional redistribution method for basic belief assignment (bba) approximation presented briefly in [16] and in more details in [17]. Some examples are given at the end of this paper to illustrate our proposed new probability transformation approach. Comparisons of our new HDSmP approach with the other well-known approaches with related analyses are also provided.

II. EVIDENCE THEORY AND PROBABILITY TRANSFORMATIONS

A. Brief introduction of evidence theory

In Dempster-Shafer theory [1], the elements in the frame of discernment (FOD) Θ are mutually exclusive. Suppose that 2^Θ represents the powerset of FOD, and one defines the function $m : 2^\Theta \to [0,1]$ as the basic belief assignment (bba), also called mass function satisfying:

$$\sum_{A \subseteq \Theta} m(A) = 1, \ m(\emptyset) = 0 \qquad (1)$$

Belief function (Bel) and plausibility function (Pl) are defined below, respectively:

$$Bel(A) = \sum_{B \subseteq A} m(B) \qquad (2)$$

$$Pl(A) = \sum_{A \cap B \neq \emptyset} m(B) \qquad (3)$$

Suppose that $m_1, m_2, ..., m_n$ are n mass functions, Dempster's rule of combination is defined in (4):

$$m(A) = \begin{cases} 0, & A = \emptyset \\ \dfrac{\sum_{\cap A_i = A} \prod_{1 \leq i \leq n} m_i(A_i)}{\sum_{\cap A_i \neq \emptyset} \prod_{1 \leq i \leq n} m_i(A_i)}, & A \neq \emptyset \end{cases} \qquad (4)$$

Dempster's rule of combination is used in DST to accomplish the fusion of bodies of evidence (BOEs). However, the final goal for decision-level information fusion is decision making. The beliefs should be transformed into probabilities, before the probability-based decision-making. Although there are also some research works on making decision directly based on belief function or bba [4], probability-based decision methods are more intuitive and have become the current trend to decide under uncertainty from approximate reasoning theories [15]. Some existing and well-known probability transformation approaches are briefly reviewed in the next section.

B. Probability transformations used in DST framework

A probability transformation (or briefly a "probabilization") is a mapping $PT : Bel_\Theta \to P_\Theta$, where Bel_Θ means the belief function defined on Θ and P_Θ represents a probability measure (in fact a probability mass function, pmf) defined on Θ. Thus the probability transformation assigns a Bayesian belief function (i.e. probability measure) to any general (i.e. non-Bayesian) belief function. It is a reason why the transformations from belief functions to probability distributions are sometimes called also Bayesian transformations.

The major probability transformation approaches used so far are:

a) Pignistic transformation

The classical pignistic probability was proposed by Smets [2] in his TBM framework which is a subjective and a non-probabilistic interpretation of DST. It extends the evidence theory to the open-world propositions and it has a range of tools including discounting and conditioning to handle belief functions. At the credal level of TBM, beliefs are entertained, combined and updated. While at the pignistic level, beliefs are used to make decisions by resorting to pignistic probability transformation (PPT). The pignistic probability obtained is always called betting commitment probability (in short, BetP). The basic idea of pignistic transformation consists of transferring the positive belief of each compound (or nonspecific) element onto the singletons involved in that compound element split by the cardinality of the proposition when working with normalized bba's.

Suppose that $\Theta = \{\theta_1, \theta_2, ..., \theta_n\}$ is the FOD. The PPT for the singletons is defined as [2]:

$$\mathrm{BetP}_m(\theta_i) = \sum_{\theta_i \in B, \ B \in 2^\Theta} \frac{m(B)}{|B|} \qquad (5)$$

PPT is designed according to an idea similar to uncertainty maximization. In PPT, masses are not assigned discriminately to different singletons involved. For information fusion, the aim is to reduce the degree of uncertainty and to gain a more consolidated and reliable decision result. High uncertainty in PPT might not be helpful for the decision. To overcome this, some typical modified probability transformation approaches were proposed which are summarized below.

b) Sudano's probabilities

Sudano [8] proposed Probability transformation proportional to Plausibilities (PrPl), Probability transformation proportional to Beliefs (PrBel), Probability transformation proportional to the normalized Plausibilities (PrNPl), Probability transformation proportional to all Plausibilities (PraPl) and Hybrid Probability transformation (PrHyb), respectively. As suggested by their names, different kinds of mappings were used. For the belief function defined on the FOD $\Theta = \{\theta_1, ..., \theta_n\}$, they are respectively defined by

$$\mathrm{PrPl}(\theta_i) = Pl(\{\theta_i\}) \cdot \sum_{Y \in 2^\Theta, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} Pl(\{\theta_j\})} \qquad (6)$$

$$\mathrm{PrBel}(\theta_i) = Bel(\{\theta_i\}) \cdot \sum_{Y \in 2^\Theta, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} Bel(\{\theta_j\})} \qquad (7)$$

$$\mathrm{PrNPl}(\theta_i) = \frac{Pl(\{\theta_i\})}{\sum_j Pl(\{\theta_j\})} \qquad (8)$$

$$\mathrm{PraPl}(\theta_i) = Bel(\{\theta_i\}) + \frac{1 - \sum_j Bel(\{\theta_j\})}{\sum_j Pl(\{\theta_j\})} \cdot Pl(\{\theta_i\}) \qquad (9)$$

$$\text{PrHyb}(\theta_i) = \text{PraPl}(\theta_i) \cdot \sum_{Y \in 2^\Theta, \theta_i \in Y} \frac{m(Y)}{\sum_{\cup_j \theta_j = Y} \text{PraPl}(\theta_j)} \quad (10)$$

c) Cuzzolin's intersection probability

From a geometric interpretation of Dempster's rule of combination, an intersection probability measure was proposed by Cuzzolin [12] from the proportional repartition of the total nonspecific mass (TNSM) for each contribution of the nonspecific masses involved.

$$\text{CuzzP}(\theta_i) = m(\{\theta_i\}) + \frac{Pl(\{\theta_i\}) - m(\{\theta_i\})}{\sum_j (Pl(\{\theta_j\}) - m(\{\theta_j\}))} \cdot \text{TNSM} \quad (11)$$

where

$$\text{TNSM} = 1 - \sum_j m(\{\theta_j\}) = \sum_{A \in 2^\Theta, |A| > 1} m(A) \quad (12)$$

d) DSmP transformation

DSmP proposed recently by Dezert and Smarandache is defined as follows:

$$\text{DSmP}_\epsilon(\theta_i) = m(\{\theta_i\})$$
$$+ (m(\{\theta_i\}) + \epsilon) \cdot \sum_{\substack{X \in 2^\Theta \\ \theta_i \subset X \\ |X| \geq 2}} \frac{m(X)}{\sum_{\substack{Y \in 2^\Theta \\ Y \subset X \\ |Y| = 1}} m(Y) + \epsilon \cdot |X|} \quad (13)$$

In DSmP, both the mass assignments and the cardinality of focal elements are used in the proportional redistribution process. The parameter of ϵ is used to adjust the effect of focal element's cardinality in the proportional redistribution, and to make DSmP defined and computable when encountering zero masses. DSmP made an improvement compared with Sudano's, Cuzzolin's and PPT formulas, in that DSmP mathematically makes a more judicious redistribution of the ignorance masses to the singletons involved and thus increases the PIC level of the resulting approximation. Moreover, DSmP works for both theories of DST and DSmT.

There are still some other definitions on modified PPT such as the iterative and self-consistent approach PrScP proposed by Sudano in [5], and a modified PrScP in [11]. Although the aforementioned probability transformation approaches are different, they are all evaluated according to the degree of uncertainty. The classical evaluation criteria for a probability transformation are the following ones:

1) Normalized Shannon Entropy

Suppose that $P(\theta)$ is a probability mass function (pmf), where $\theta \in \Theta$, $|\Theta| = N$ and the $|\Theta|$ represents the cardinality of the FOD Θ. An evaluation criterion for the pmf obtained from different probability transformation is as follows [12]:

$$\text{E}_\text{H} = \frac{- \sum_{\theta \in \Theta} P(\theta) \log_2(P(\theta))}{\log_2 N} \quad (14)$$

i.e., the ratio of Shannon entropy and the maximum of Shannon entropy for $\{P(\theta)|\theta \in \Theta\}, |\Theta| = N$. Clearly E_H is normalized. The larger E_H is, the larger the degree of uncertainty is. The smaller E_H is, the smaller the degree of uncertainty is. When $\text{E}_\text{H} = 0$, one hypothesis will have probability 1 and the rest with zero probabilities. Therefore the agent or system can make decision without error. When $\text{E}_\text{H} = 1$, it is impossible to make a correct decision, because $P(\theta)$, for all $\theta \in \Theta$ are equal.

2) Probabilistic Information Content

Probabilistic Information Content (PIC) criterion [5] is an essential measure in any threshold-driven automated decision system. The PIC value of a pmf obtained from a probability transformation indicates the level of the total knowledge one has to draw a correct decision.

$$\text{PIC}(P) = 1 + \frac{1}{\log_2 N} \cdot \sum_{\theta \in \Theta} P(\theta) \log_2(P(\theta)) \quad (15)$$

Obviously, PIC $= 1 - \text{E}_\text{H}$. The PIC is the dual of the normalized Shannon entropy. A PIC value of zero indicates that the knowledge to take a correct decision does not exist (all hypotheses have equal probabilities, i.e., one has the maximal entropy).

Less uncertainty means that the corresponding probability transformation result is better to help to take a decision. According to such a simple and basic idea, the probability transformation approach should attempt to enlarge the belief differences among all the propositions and thus to achieve a more reliable decision result.

III. THE HIERARCHICAL DSmP TRANSFORMATION

In this paper, we propose a novel probability transformation approach called hierarchical DSmP (HDSmP), which provides a new way to reduce step by step the mass committed to uncertainties until to obtain an approximate measure of subjective probability, i.e. a so-called Bayesian bba in [1]. It must be noticed that this procedure can be stopped at any step in the process and thus it allows to reduce the number of focal elements in a given bba in a consistent manner to diminish the size of the core of a bba and thus reduce the complexity (if needed) when applying also some complex rules of combinations. We present here the general principle of hierarchical and proportional reduction of uncertainties in order to finally obtain a Bayesian bba. The principle of redistribution of uncertainty to more specific elements of the core at any given step of the process follows the proportional redistribution already proposed in the (non hierarchical) DSmP transformation proposed recently in [3].

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 $X(2)$ denotes 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 s involving elements Y and X of the powerset is defined as (for $X \neq \emptyset$ and $Y \neq \emptyset$)

$$R_s(Y, X) \triangleq \frac{m(Y) + \epsilon \cdot |X|}{\sum_{\substack{Y \subset X \\ |X| - |Y| = s}} (m(Y) + \epsilon \cdot |X|)} \quad (16)$$

where ϵ is a small positive number introduced here to deal with particular cases where $\sum_{\substack{Y \subset X \\ |X| - |Y| = s}} m(Y) = 0$.
In HDSmP, we just need to use the proportional redistribution factors of width $n = 1$, and so we will just denote $R(Y, X) \triangleq R_1(Y, X)$ by convention.

The HDSmP transformation 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. Mathematically, HDSmP is defined for any $X(1) \in \Theta$, i.e. any $\theta_i \in \Theta$ by

$$HDSmP(X(1)) = m(X(1)) +$$
$$\sum_{\substack{X(2) \supset X(1) \\ X(1), X(2) \in 2^\Theta}} [m_h(X(2)) \cdot R(X(1), X(2))] \quad (17)$$

where the "hierarchical" masses $m_h(.)$ are recursively (backward) computed as follows:

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

$$m_h(A) = m(A), \forall |A| < n - 1 \quad (18)$$

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

$$m_h(A) = m(A), \forall |A| < n - 2$$
$$\vdots \quad (19)$$

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

$$m_h(A) = m(A), \forall |A| < 2 \quad (20)$$

Actually, it is worth to note that $X(n)$ is in fact unique and it corresponds only to the full ignorance $\theta_1 \cup \theta_2 \cup \ldots \cup \theta_n$. Therefore, the expression of $m_h(X(n - 1))$ in Eq. (18) just simplifies as

$$m_h(X(n-1)) = m(X(n-1)) + m(X(n)) \cdot R(X(n-1), X(n))$$

Because of the full proportional redistribution of the masses of uncertainties to the elements least specific involved in these uncertainties, no mass of belief is lost during the step by step hierarchical process and thus one gets finally a Bayesian bba satisfying $\sum_{X(1) \in 2^\Theta} HDSmP(X(1)) = 1$.

IV. EXAMPLES

In this section we show in details how HDSmP can be applied on very simple different examples. So let's examine the three following examples based on a simple 3D frame of discernment $\Theta = \{\theta_1, \theta_2, \theta_3\}$ satisfying Shafer's model.

A. Example 1

Let's consider the following bba:

$$m(\theta_1) = 0.10, \quad m(\theta_2) = 0.17, \quad m(\theta_3) = 0.03,$$
$$m(\theta_1 \cup \theta_2) = 0.15, \quad m(\theta_1 \cup \theta_3) = 0.20,$$
$$m(\theta_2 \cup \theta_3) = 0.05, \quad m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30.$$

We apply HDSmP with $\epsilon = 0$ in this example because there is no mass of belief equal to zero. It can be verified that the result obtained with a small positive ϵ parameter remains (as expected) numerically very close to the result obtained with $\epsilon = 0$. This verification is left to the reader.

The first step of HDSmP consists in redistributing back $m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30$ committed to the full ignorance to the elements $\theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_2 \cup \theta_3$ only, because these elements are the only elements of cardinality 2 that are included in $\theta_1 \cup \theta_2 \cup \theta_3$. Applying the Eq. (18) with $n = 3$, one gets when $X(2) = \theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_1 \cup \theta_2$ the following masses

$$m_h(\theta_1 \cup \theta_2) = m(\theta_1 \cup \theta_2) + m(X(3)) \cdot R(\theta_1 \cup \theta_2, X(3))$$
$$= 0.15 + (0.3 \cdot 0.375) = 0.2625$$

because $R(\theta_1 \cup \theta_2, X(3)) = \frac{0.15}{0.15 + 0.20 + 0.05} = 0.375$.
Similarly, one gets

$$m_h(\theta_1 \cup \theta_3) = m(\theta_1 \cup \theta_3) + m(X(3)) \cdot R(\theta_1 \cup \theta_3, X(3))$$
$$= 0.20 + (0.3 \cdot 0.5) = 0.35$$

because $R(\theta_1 \cup \theta_3, X(3)) = \frac{0.20}{0.15 + 0.20 + 0.05} = 0.5$, and also

$$m_h(\theta_2 \cup \theta_3) = m(\theta_2 \cup \theta_3) + m(X(3)) \cdot R(\theta_2 \cup \theta_3, X(3))$$
$$= 0.05 + (0.3 \cdot 0.125) = 0.0875$$

because $R(\theta_2 \cup \theta_3, X(3)) = \frac{0.05}{0.15 + 0.20 + 0.05} = 0.125$.

Now, we go to the next step of HDSmP and one needs to redistribute the masses of partial ignorances $X(2)$ corresponding to $\theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_2 \cup \theta_3$ back to the singleton elements $X(1)$ corresponding to θ_1, θ_2 and θ_3. We use directly HDSmP in Eq. (17) for doing this as follows:

$$HDSmP(\theta_1) = m(\theta_1) + m_h(\theta_1 \cup \theta_2) \cdot R(\theta_1, \theta_1 \cup \theta_2)$$
$$+ m_h(\theta_1 \cup \theta_3) \cdot R(\theta_1, \theta_1 \cup \theta_3)$$
$$\approx 0.10 + (0.2625 \cdot 0.3703) + (0.35 \cdot 0.7692)$$
$$= 0.10 + 0.0972 + 0.2692 = 0.4664$$

because

$$R(\theta_1, \theta_1 \cup \theta_2) = \frac{0.10}{0.10 + 0.17} \approx 0.3703$$

$$R(\theta_1, \theta_1 \cup \theta_3) = \frac{0.10}{0.10 + 0.03} \approx 0.7692$$

Similarly, one gets

$$HDSmP(\theta_2) = m(\theta_2) + m_h(\theta_1 \cup \theta_2) \cdot R(\theta_2, \theta_1 \cup \theta_2)$$
$$+ m_h(\theta_2 \cup \theta_3) \cdot R(\theta_2, \theta_2 \cup \theta_3)$$
$$\approx 0.10 + (0.2625 \cdot 0.6297) + (0.0875 \cdot 0.85)$$
$$= 0.17 + 0.1653 + 0.0744 = 0.4097$$

because

$$R(\theta_2, \theta_1 \cup \theta_2) = \frac{0.17}{0.10 + 0.17} \approx 0.6297$$

$$R(\theta_2, \theta_2 \cup \theta_3) = \frac{0.17}{0.17 + 0.03} = 0.85$$

and also

$$HDSmP(\theta_3) = m(\theta_3) + m_h(\theta_1 \cup \theta_3) \cdot R(\theta_3, \theta_1 \cup \theta_3)$$
$$+ m_h(\theta_2 \cup \theta_3) \cdot R(\theta_3, \theta_2 \cup \theta_3)$$
$$\approx 0.03 + (0.35 \cdot 0.2307) + (0.0875 \cdot 0.15)$$
$$= 0.03 + 0.0808 + 0.0131 = 0.1239$$

because

$$R(\theta_3, \theta_1 \cup \theta_3) = \frac{0.03}{0.10 + 0.03} \approx 0.2307$$

$$R(\theta_3, \theta_2 \cup \theta_3) = \frac{0.03}{0.17 + 0.03} = 0.15$$

Hence, the final result of HDSmP transformation is:

$$HDSmP(\theta_1) = 0.4664, \quad HDSmP(\theta_2) = 0.4097,$$
$$HDSmP(\theta_3) = 0.1239.$$

and we can easily verify that

$$HDSmP(\theta_1) + HDSmP(\theta_2) + HDSmP(\theta_3) = 1.$$
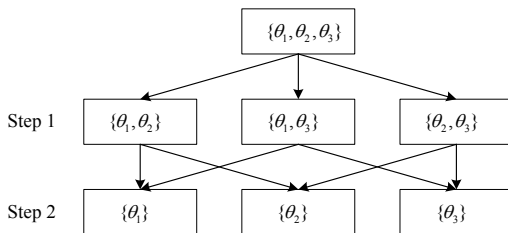
The procedure can be illustrated in Fig. 1 below.

Figure 1. Illustration of Example 1

Table I
EXPERIMENTAL RESULTS FOR EXAMPLE 1.

Approaches	Propositions			E_H
	θ_1	θ_2	θ_3	
BetP	0.3750	0.3700	0.2550	0.9868
PrPl	0.4045	0.3681	0.2274	0.9747
PrBel	0.4094	0.4769	0.1137	0.8792
DSmP_0	0.4094	0.4769	0.1137	0.8792
DSmP_0.001	0.4094	0.4769	0.1137	0.8792
HDSmP_0	0.4664	0.4097	0.1239	0.8921
HDSmP_0.001	0.4664	0.4097	0.1239	0.8921

The classical DSmP transformation [3] and the other transformations (BetP [2], PrBel and PrPl [8]) are compared with HDSmP for this example in Table I. It can be seen in Table I that the normalized entropy E_H of HDSmP is relatively small but not too small among all the probability transformations used. In fact it is normal that the entropy drawn form HDSmP is a bit bigger than the entropy drawn from DSmP, because there is a "dilution" of uncertainty in the step-by-step redistribution, whereas such dilution of uncertainty is absent in the direct DSmP transformation.

B. Example 2

Let's consider the following bba:

$$m(\theta_1) = 0, \quad m(\theta_2) = 0.17, \quad m(\theta_3) = 0.13,$$
$$m(\theta_1 \cup \theta_2) = 0.20, \quad m(\theta_1 \cup \theta_3) = 0.20,$$
$$m(\theta_2 \cup \theta_3) = 0, \quad m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30$$

The first step of HDSmP consists in redistributing back $m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30$ committed to the full ignorance to the elements $\theta_1 \cup \theta_2$, and $\theta_1 \cup \theta_3$ only, because these elements are the only elements of cardinality 2 that are included in $\theta_1 \cup \theta_2 \cup \theta_3$. Applying the Eq. (18) with $n = 3$, one gets when $X(2) = \theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_1 \cup \theta_2$ the following masses

$$m_h(\theta_1 \cup \theta_2) = m(\theta_1 \cup \theta_2) + m(X(3)) \cdot R(\theta_1 \cup \theta_2, X(3))$$
$$= 0.20 + (0.3 \cdot 0.5) = 0.35$$

because $R(\theta_1 \cup \theta_2, X(3)) = \frac{0.20}{0.20 + 0.20 + 0.00} = 0.5$.
Similarly, one gets

$$m_h(\theta_1 \cup \theta_3) = m(\theta_1 \cup \theta_3) + m(X(3)) \cdot R(\theta_1 \cup \theta_3, X(3))$$
$$= 0.20 + (0.3 \cdot 0.5) = 0.35$$

because $R(\theta_1 \cup \theta_3, X(3)) = \frac{0.20}{0.20 + 0.20 + 0.00} = 0.5$, and also

$$m_h(\theta_2 \cup \theta_3) = m(\theta_2 \cup \theta_3) + m(X(3)) \cdot R(\theta_2 \cup \theta_3, X(3))$$
$$= 0.00 + (0.3 \cdot 0.0) = 0$$

because $R(\theta_2 \cup \theta_3, X(3)) = \frac{0.0}{0.20 + 0.20 + 0.00} = 0$.

Now, we go to the next and last step of HDSmP principle, and one needs to redistribute the masses of partial ignorances $X(2)$ corresponding to $\theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_2 \cup \theta_3$ back to the singleton elements $X(1)$ corresponding to θ_1, θ_2 and θ_3.

We use directly HDSmP in Eq. (17) for doing this as follows:

$$HDSmP(\theta_1) = m(\theta_1) + m_h(\theta_1 \cup \theta_2) \cdot R(\theta_1, \theta_1 \cup \theta_2)$$
$$+ m_h(\theta_1 \cup \theta_3) \cdot R(\theta_1, \theta_1 \cup \theta_3)$$
$$\approx 0.00 + (0.35 \cdot 0.00) + (0.35 \cdot 0.00)$$
$$= 0.00 + 0.00 + 0.00 = 0$$

because

$$R(\theta_1, \theta_1 \cup \theta_2) = \frac{0.00}{0.00 + 0.17} = 0.00$$

$$R(\theta_1, \theta_1 \cup \theta_3) = \frac{0.00}{0.00 + 0.13} = 0.00$$

Similarly, one gets

$$HDSmP(\theta_2) = m(\theta_2) + m_h(\theta_1 \cup \theta_2) \cdot R(\theta_2, \theta_1 \cup \theta_2)$$
$$+ m_h(\theta_2 \cup \theta_3) \cdot R(\theta_2, \theta_2 \cup \theta_3)$$
$$\approx 0.17 + (0.35 \cdot 1) + (0.00 \cdot 0.5667)$$
$$= 0.17 + 0.35 + 0.00 = 0.52$$

because

$$R(\theta_2, \theta_1 \cup \theta_2) = \frac{0.17}{0.00 + 0.17} = 1$$

$$R(\theta_2, \theta_2 \cup \theta_3) = \frac{0.17}{0.17 + 0.13} \approx 0.5667$$

and also

$$HDSmP(\theta_3) = m(\theta_3) + m_h(\theta_1 \cup \theta_3) \cdot R(\theta_3, \theta_1 \cup \theta_3)$$
$$+ m_h(\theta_2 \cup \theta_3) \cdot R(\theta_3, \theta_2 \cup \theta_3)$$
$$\approx 0.13 + (0.35 \cdot 1) + (0.00 \cdot 0.4333)$$
$$= 0.13 + 0.35 + 0.00 = 0.48$$

because

$$R(\theta_3, \theta_1 \cup \theta_3) = \frac{0.13}{0.13 + 0.00} = 1$$

$$R(\theta_3, \theta_2 \cup \theta_3) = \frac{0.13}{0.17 + 0.13} \approx 0.4333$$

Hence, the final result of HDSmP transformation is:

$$HDSmP(\theta_1) = 0.4664, \qquad HDSmP(\theta_2) = 0.4097,$$
$$HDSmP(\theta_3) = 0.1239.$$

and we can easily verify that

$$HDSmP(\theta_1) + HDSmP(\theta_2) + HDSmP(\theta_3) = 1.$$

The HDSmP procedure of Example 2 with $\epsilon = 0$ is Fig. 2. The HDSmP procedure of Example 2 with $\epsilon > 0$ is the same as that illustrated in Fig. 1. When one takes $\epsilon > 0$, there exist masses redistributed to $\{\theta_2 \cup \theta_3\}$. If one takes $\epsilon = 0$, there is no mass edistributed to $\{\theta_2 \cup \theta_3\}$. That's the difference between Fig. 1 and Fig. 2.

Let's suppose that one takes $\epsilon = 0.001$, then the HDSmP calculation procedure is as follows:
• Step 1: The first step of HDSmP consists in distributing back $m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30$ committed to the full ignorance to the elements $\theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_2 \cup \theta_3$. Applying the formula
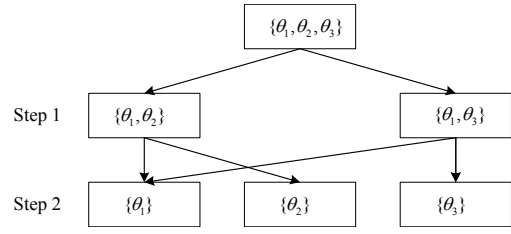

Figure 2. Illustration of Example 2.

(III) with $n = 3$, one gets when $X(2) = \theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_1 \cup \theta_2$ the following masses

$$m_h(\theta_1 \cup \theta_2) = m(\theta_1 \cup \theta_2) + m(X(3)) \cdot R(\theta_1 \cup \theta_2, X(3))$$
$$= 0.20 + (0.3 \cdot 0.4963) = 0.3489$$

because

$$R(\theta_1 \cup \theta_2, X(3)) = \frac{0.20 + 0.001 \cdot 3}{(0.20 + 0.001 \cdot 3) \cdot 2 + (0.00 + 0.001 \cdot 3)}$$
$$= 0.4963$$

$$m_h(\theta_1 \cup \theta_3) = m(\theta_1 \cup \theta_3) + m(X(3)) \cdot R(\theta_1 \cup \theta_3, X(3))$$
$$= 0.20 + (0.3 \cdot 0.4963) = 0.3489$$

because

$$R(\theta_1 \cup \theta_2, X(3)) = \frac{0.20 + 0.001 \cdot 3}{(0.20 + 0.001 \cdot 3) \cdot 2 + (0.00 + 0.001 \cdot 3)}$$
$$= 0.4963$$

$$m_h(\theta_2 \cup \theta_3) = m(\theta_2 \cup \theta_3) + m(X(3)) \cdot R(\theta_2 \cup \theta_3, X(3))$$
$$= 0.00 + (0.3 \cdot 0.0073) = 0.0022$$

because

$$R(\theta_2 \cup \theta_3, X(3)) = \frac{0.001 \cdot 3}{(0.20 + 0.001 \cdot 3) \cdot 2 + (0.00 + 0.001 \cdot 3)}$$
$$= 0.0073$$

• Next step: one needs to redistribute the masses of partial ignorances $X(2)$ corresponding to $\theta_1 \cup \theta_2$, $\theta_1 \cup \theta_3$ and $\theta_2 \cup \theta_3$ back to the singleton elements $X(1)$ corresponding to θ_1, θ_2 and θ_3. We use directly HDSmP in Eq. (17) for doing this as follows:

$$HDSmP(\theta_1) = m(\theta_1) + m_h(\theta_1 \cup \theta_2) \cdot R(\theta_1, \theta_1 \cup \theta_2)$$
$$+ m_h(\theta_1 \cup \theta_3) \cdot R(\theta_1, \theta_1 \cup \theta_3)$$
$$\approx 0.00 + (0.3489 \cdot 0.0115) + (0.3489 \cdot 0.0149)$$
$$= 0.00 + 0.0040 + 0.0052 = 0.0092$$

because

$$R(\theta_1, \theta_1 \cup \theta_2) = \frac{0.00 + 0.001 \cdot 2}{(0.00 + 0.001 \cdot 2) + (0.17 + 0.001 \cdot 2)}$$
$$= 0.0115$$

$$R(\theta_1, \theta_1 \cup \theta_3) = \frac{0.00 + 0.001 \cdot 2}{(0.00 + 0.001 \cdot 2) + (0.13 + 0.001 \cdot 2)}$$
$$= 0.0149$$

Similarly, one gets

$$HDSmP(\theta_2) = m(\theta_2) + m_h(\theta_1 \cup \theta_2) \cdot R(\theta_2, \theta_1 \cup \theta_2)$$
$$+ m_h(\theta_2 \cup \theta_3) \cdot R(\theta_2, \theta_2 \cup \theta_3)$$
$$\approx 0.17 + (0.3489 \cdot 0.9885) + (0.0022 \cdot 0.5658)$$
$$= 0.17 + 0.3449 + 0.0012 = 0.5161$$

because

$$R(\theta_2, \theta_1 \cup \theta_2) = \frac{0.17 + 0.001 \cdot 2}{(0.00 + 0.001 \cdot 2) + (0.17 + 0.001 \cdot 2)}$$
$$= 0.9885$$

$$R(\theta_2, \theta_2 \cup \theta_3) = \frac{0.17 + 0.001 \cdot 2}{(0.17 + 0.001 \cdot 2) + (0.13 + 0.001 \cdot 2)}$$
$$\approx 0.5658$$

and also

$$HDSmP(\theta_3) = m(\theta_3) + m_h(\theta_1 \cup \theta_3) \cdot R(\theta_3, \theta_1 \cup \theta_3)$$
$$+ m_h(\theta_2 \cup \theta_3) \cdot R(\theta_3, \theta_2 \cup \theta_3)$$
$$\approx 0.13 + (0.3489 \cdot 0.9851) + (0.0022 \cdot 0.4342)$$
$$= 0.13 + 0.3437 + 0.0009 = 0.4746$$

because

$$R(\theta_3, \theta_1 \cup \theta_3) = \frac{0.13 + 0.001 \cdot 2}{(0.13 + 0.001 \cdot 2) + (0.00 + 0.001 \cdot 2)}$$
$$= 0.9851$$

$$R(\theta_3, \theta_2 \cup \theta_3) = \frac{0.13 + 0.001 \cdot 2}{(0.17 + 0.001 \cdot 2) + (0.13 + 0.001 \cdot 2)}$$
$$\approx 0.4342$$

Hence, the final result of HDSmP transformation is:

$$HDSmP(\theta_1) = 0.0092, \quad HDSmP(\theta_2) = 0.5161,$$
$$HDSmP(\theta_3) = 0.4746.$$

and we can easily verify that

$$HDSmP(\theta_1) + HDSmP(\theta_2) + HDSmP(\theta_3) = 1.$$

We also calculate some other probability transformations and the results are listed in Table II.

Table II
EXPERIMENTAL RESULTS FOR EXAMPLE 2.

Approaches	Propositions			E_H
	θ_1	θ_2	θ_3	
BetP	0.3000	0.3700	0.3300	0.9966
PrPl	0.3125	0.3683	0.3192	0.9975
PrBel	NaN	NaN	NaN	NaN
DSmP_0	0.0000	0.5400	0.4600	0.6280
DSmP_0.001	0.0037	0.5381	0.4582	0.6479
HDSmP_0	0.0000	0.5200	0.4800	0.6302
HDSmP_0.001	0.0092	0.5161	0.4746	0.6720

It can be seen in Table II that the normalized entropy E_H of HDSmP is relatively small but not too small among all the probability transformations used.

C. Example 3

Let's consider the following bba:

$$m(\theta_1) = 0, \quad m(\theta_2) = 0, \quad m(\theta_3) = 0.70,$$
$$m(\theta_1 \cup \theta_2) = 0, \quad m(\theta_1 \cup \theta_3) = 0,$$
$$m(\theta_2 \cup \theta_3) = 0, \quad m(\theta_1 \cup \theta_2 \cup \theta_3) = 0.30$$

In this example, the mass assignments for all the focal elements with cardinality size 2 equal to zero. For HDSmP, when $\epsilon > 0$, $m(\theta_2 \cup \theta_3)$ will be divided equally and redistributed to $\{\theta_1 \cup \theta_2\}$, $\{\theta_1 \cup \theta_3\}$ and $\{\theta_2 \cup \theta_3\}$. Because the ratios are

$$R(\theta_1 \cup \theta_2, X(3)) = R(\theta_1 \cup \theta_3, X(3)) = R(\theta_2 \cup \theta_3, X(3))$$
$$= \frac{0.00 + 0.001 \cdot 3}{(0.00 + 0.001 \cdot 3) \cdot 3} = 0.3333$$

One sees that with the parameter $\epsilon = 0$, HDSmP cannot be computed (division by zero) and that is why it is necessary to use $\epsilon > 0$ in such particular case. The results of HDSmP and other probability transformations are listed in Table III.

Table III
EXPERIMENTAL RESULTS FOR EXAMPLE 3.

Approaches	Propositions			E_H
	θ_1	θ_2	θ_3	
BetP	0.1000	0.1000	0.8000	0.5871
PrPl	0.0562	0.0562	0.8876	0.3911
PrBel	NaN	NaN	NaN	NaN
DSmP_0	0.0000	0.0000	1.0000	0.0000
DSmP_0.001	0.0004	0.0004	0.0092	0.0065
HDSmP_0	NaN	NaN	NaN	NaN
HDSmP_0.001	0.0503	0.0503	0.8994	0.3606

It can be seen in Table III that the normalized entropy E_H of HDSmP is relatively small but not the smallest among all the probability transformations used. Naturally, and as already pointed out, HDSmP$_{\epsilon=0}$ cannot be computed in such example because of division by zero. But with the use of the parameter $\epsilon = 0.001$, the mass of $m(\theta_1 \cup \theta_2 \cup \theta_3)$ becomes equally divided and redistributed to the focal elements with cardinality of 2. This justify the necessity of the use of parameter $\epsilon > 0$ in some particular cases when there exist masses equal to zero.

D. Example 4 (vacuous bba)

Let's consider the following particular bba, called the vacuous bba since it represents a fully ignorant source of evidence:

$$m(\theta_1) = 0, \quad m(\theta_2) = 0, \quad m(\theta_3) = 0,$$
$$m(\theta_1 \cup \theta_2) = 0, \quad m(\theta_1 \cup \theta_3) = 0,$$
$$m(\theta_2 \cup \theta_3) = 0, \quad m(\theta_1 \cup \theta_2 \cup \theta_3) = 1$$

In this example, the mass assignments for all the focal elements with cardinality less than 3 equal to zero. For HDSmP, when $\epsilon > 0$, $m(\theta_1 \cup \theta_2 \cup \theta_3)$ will be divided equally and redistributed to $\{\theta_1 \cup \theta_2\}$, $\{\theta_1 \cup \theta_3\}$ and $\{\theta_2 \cup \theta_3\}$. Similarly, the mass assignments for focal elements with cardinality of 2 (partial ignorances) obtained at the intermediate step will be divided equally and redistributed to singletons included in them. This redistribution is possible for the existence of $\epsilon > 0$ in HDSmP formulas. HDSmP cannot be applied and computed

in such example if one takes $\epsilon = 0$, and that is why one needs to use $\epsilon > 0$ here. The results of HDSmP and other probability transformations are listed in Table IV.

Table IV
EXPERIMENTAL RESULTS FOR EXAMPLE 4.

Approaches	Propositions			E_H
	θ_1	θ_2	θ_3	
BetP	0.3333	0.3333	0.3333	1.0000
PrPl	0.3333	0.3333	0.3333	1.0000
PrBel	NaN	NaN	NaN	NaN
$DSmP_{-0}$	NaN	NaN	NaN	NaN
$DSmP_{-0.001}$	0.3333	0.3333	0.3333	1.0000
$HDSmP_{-0}$	NaN	NaN	NaN	NaN
$HDSmP_{-0.001}$	0.3333	0.3333	0.3333	1.0000

It can be seen in Tables I – IV that the normalized entropy E_H of HDSmP is always moderate among the other probability transformations it is compared with, and it is normal to get an entropy value with HDSmP bigger than with DSmP because of dilution of uncertainty through the procedure of HDSmP. We have already shown that the entropy criteria is not enough in fact to evaluate the quality a probability transformation [14], and always a compromise must be found between entropy level and numerical robustness of the transformation. Although the entropy should be as small as possible for decision-making, exaggerate small entropy is not always preferred. Because of the way the mass of (partial) ignorances is proportionally redistributed, it is clear that if the mass assignment for a singleton equals to zero in the original bba, then after applying DSmP or HDSmP transformations this mass is unchanged and is kept to zero. This behavior may appear a bit intuitively surprising at the first glance specially if some masses of partial ignorances including this singleton are not equal to zero. This behavior is however normal in the spirit of proportional redistribution because one wants to reduce the PIC value so that if one has no strong support (belief) in a singleton in the original bba, we expect also to have no strong support in this singleton after the transformation is applied which makes perfectly sense. Of course if such behavior is considered as too optimistic or not acceptable because it appears too risky in some applications, it is always possible to choose another transformation instead. The final choice is always left in the hands of the user, or the fusion system designer.

V. CONCLUSIONS

Probability transformation is very crucial for decision-making in evidence theory. In this paper a novel interesting and useful hierarchical probability transformation approach called HDSmP has been proposed, and HDSmP always provides a moderate value of entropy which is necessary for an easier and reliable decision-making support. Unfortunately the PIC (or entropy) level is not the unique useful criterion to evaluate the quality of a probability transformation in general. At least the numerical robustness of the method is also important and must be considered seriously as already shown in our previous works. Therefore, to evaluate any probability transformation more efficiently and to outperform existing transformations (including DSmP and HDSmP) a more general comprehensive evaluation criteria need to be found. The search for such a criteria is under investigations.

REFERENCES

[1] G. Shafer, *A Mathematical Theory of Evidence*, Princeton University, Princeton, 1976.
[2] P. Smets, R. Kennes, *The transferable belief model*, Artificial Intelligence, Vol 66, No. 2, pp. 191-234, 1994.
[3] 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.
[4] W.H. Lv, *Decision-making rules based on belief interval with D-S evidence theory*, book section in Fuzzy Information and Engineering, pp. 619-627, 2007.
[5] J. Sudano, *Pignistic probability transforms for mixes of low- and high-probability events*, in Proc. of Int. Conf. on Information Fusion 2001, Montreal, Canada, , TUB3, pp. 23–27, August, 2001.
[6] J. Sudano, *The system probability information content (PIC) relationship to contributing components, combining independent multisource beliefs, hybrid and pedigree pignistic probabilities*, in Proc. of Int. Conf. on Information Fusion 2002, Annapolis, USA, Vol 2, pp. 1277–1283, July 2002.
[7] J. Sudano, *Equivalence between belief theories and naive Bayesian fusion for systems with independent evidential data - Part I, The theory*, in Proc. of Int. Conf. on Information Fusion 2003, Cairns, Australia, Vol. 2, pp. 1239–1243, July 2003.
[8] J. Sudano, *Yet another paradigm illustrating evidence fusion (YAPIEF)*, in Proc. of Int. Conf. on Information Fusion 2006, Florence, pp. 1–7, July 2006.
[9] J. Dezert, F. Smarandache, *A new probabilistic transformation of belief mass assignment*, in Int. Conf. on Information Fusion 2008, Germany, Cologne, pp. 1-8, June 30th – July 3rd, 2008.
[10] J. Sudano, *Belief fusion, pignistic probabilities, and information content in fusing tracking attributes*, in Proc. of Radar Conference, Volume, Issue, 26-29 , pp. 218 – 224, 2004.
[11] Y. Deng, W. Jiang, Q. Li, *Probability transformation in transferable belief model based on fractal theory* (in Chinese), in Proc. of Chinese Conf. on Information Fusion 2009, Yantai, China, pp. 10-13, 79, Nov. 2009
[12] W. Pan, H.J. Yang, *New methods of transforming belief functions to pignistic probability functions in evidence theory*, Int. Workshop on Intelligent Systems and Applications, pp. 1–5, Wuhan, China, May, 2009.
[13] F. Cuzzolin, *On the properties of the Intersection probability*, submitted to the Annals of Mathematics and AI, Feb. 2007.
[14] D.Q. Han, J. Dezert, C.Z. Han, Y. Yang, *Is Entropy Enough to Evaluate the Probability Transformation Approach of Belief Function?*, in Proc. of 13th Int. Conf. on Information Fusion Conference, pp. 1–7, Edinburgh, UK, July 26-29th, 2010.
[15] P. Smets, *Decision making in the TBM: the necessity of the pignistic transformation*, in Int. Journal of Approximate Reasoning, Vol 38, pp. 133–147, 2005.
[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] J. Dezert, D.Q. Han, Z.G. Liu, J.-M. Tacnet, *Hierarchical Proportional Redistribution principle for uncertainty reduction and bba approximation*, in Proceedings of the 10th World Congress on Intelligent Control and Automation (WCICA 2012), Beijing, China, July 6-8, 2012.