On open problems associated with conditioning in the Dempster-Shafer belief function theory

Radim JIROUŠEK[1], Václav KRATOCHVÍL[1] and Prakash P. SHENOY[2]

[1] *Institute of Information Theory and Automation*
Czech Academy of Sciences
Pod Vodárenskou věží 4, 18200 Prague, Czech Republic
{radim,velorex}@utia.cas.cz
[2] *University of Kansas School of Business, Lawrence, KS 66045 USA*
pshenoy@ku.edu

Abstract

As in probability theory, graphical and compositional models in the Dempster-Shafer (D-S) belief function theory handle multidimensional belief functions applied to support inference for practical problems. Both types of models represent multidimensional belief functions using their low-dimensional marginals. In the case of graphical models, these marginals are usually conditionals; for compositional models, they are unconditional. Nevertheless, one must introduce some conditioning to compose unconditional belief functions and avoid double-counting knowledge. Thus, conditioning is crucial in processing multidimensional compositional models for belief functions.

This paper summarizes some important open problems, the solution of which should enable a trouble-free design of computational processes employing D-S belief functions in AI. For some of them, we discuss possible solutions. The problems considered in this paper are of two types. There are still some gaps that should be filled to get a mathematically consistent uncertainty theory. Other problems concern the computational tractability of procedures arising from the super-exponential growth of the space and time complexity of the designed algorithms.

Keywords: belief functions, conditioning, composition, conditional independence.

1 Introduction

By compositional models, we mean multidimensional belief functions constructed from low-dimensional belief functions using some standard composition operator. Such models are much less space-demanding than general belief functions, and the computations with them should be faster and sometimes even better justifiable.

When considering probabilistic compositional models, we mean multidimensional probability distributions composed from their low-dimensional marginals. Similarly, within the framework of D-S belief functions, we consider multidimensional basic probability assignments (BPAs) assembled from a system of low-dimensional (marginal) BPAs. The beneficial effect of their use is apparent. The cardinality of state spaces, for which BPAs are defined, grows super-exponentially with the number of variables. It reflects in the computational complexity of some procedures, even if we have BPAs with few focal elements (BPAs representable by a small number of parameters). Namely, some procedures must go through all states, regardless of the number of focal elements defining the BPAs.

Considering BPAs representable by a "reasonable" number of parameters means we cannot handle all possible belief functions. We can process only the belief functions for which a system of conditional independence relations holds. This is similar to the framework of probability theory. In this paper, we consider the notion of conditional independence relation introduced in [1], though many others were introduced in the literature, as, e.g., [2, 3, 4]. The other notion closely connected with all the methods the authors know for efficiently representing multidimensional models is the notion of conditional BPAs. Without conditional BPAs, one could not set up directed graphical belief function models. Without conditioning, we would not be able to define a composition operator, and we would not be able to construct compositional models. Thus, after Section 2, where the notation and basic notions of D-S belief functions are stated, in Section 3, we present the open problems connected with conditioning. Section 4 presents open problems associated with applying compositional models to inference.

2 Belief Function Notation

This paper uses the same notation as in our paper presented at ISIPTA'23 [5]. X, Y, ... denote discrete (finite-valued) variables. Lower-case characters r, s, t, ... denote the sets of variables. Ω_X, $\Omega_Y, \ldots$ denote the state spaces of the corresponding variables. For a set of variables r, the corresponding state space is a Cartesian product $\Omega_r = \times_{X \in r} \Omega_X$. 2^{Ω_r} will denote the set of all subsets of Ω_r.

A *basic probability assignment* (BPA) for r is a mapping $m : 2^{\Omega_r} \to [0,1]$, such that $\sum_{\mathsf{a} \subseteq \Omega_r} m(\mathsf{a}) = 1$ and $m(\emptyset) = 0$. We often call it a *joint* BPA to highlight that it is defined for a group of variables r. We say that $\mathsf{a} \subseteq \Omega_r$ is a *focal element* of m if $m(\mathsf{a}) > 0$. A BPA with only one focal element is called *deterministic*. ι_r denotes the deterministic BPA for r, the focal element of which is the entire state space: $\iota_r(\Omega_r) = 1$. Since ι_r represents total ignorance, it is called *vacuous*. BPA m is said to be Bayesian if all its focal elements are singletons: $(m(\mathsf{a}) > 0 \Rightarrow |\mathsf{a}| = 1)$.

A BPA m for r can also be specified by a corresponding *plausibility function*, *belief function* (BEL), and *commonality function* (CF) [6]. These functions are also mappings $2^{\Omega_r} \to [0,1]$. The latter two can be derived from BPA m as follows:

$$Bel_m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_r :\, \mathsf{b} \subseteq \mathsf{a}} m(\mathsf{b}), \tag{1}$$

$$Q_m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_r :\, \mathsf{b} \supseteq \mathsf{a}} m(\mathsf{b}). \tag{2}$$

These representations are equivalent; we can uniquely compute the others when one of them is given:

$$m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \mathsf{a}} (-1)^{|\mathsf{a} \setminus \mathsf{b}|} Bel_m(\mathsf{b}), \tag{3}$$

$$m(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_r :\, \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q_m(\mathsf{b}). \tag{4}$$

Based on the requirement of non-negativity and normality of BPAs, and on Eq. (4), it follows that function $Q : 2^{\Omega}_r \to [0,1]$ is a CF for r iff

$$Q(\emptyset) = 1, \tag{5}$$

$$\sum_{\mathsf{b} \subseteq \Omega_r :\, \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} Q(\mathsf{b}) \geq 0, \qquad \text{for all } \mathsf{a} \subseteq \Omega_r, \text{and} \tag{6}$$

$$\sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_r} (-1)^{|\mathsf{a}|+1} Q(\mathsf{a}) = 1. \tag{7}$$

It follows from Eq. (2) that a CF is non-increasing in the sense that

$$\mathsf{a} \subseteq \mathsf{b} \implies Q(\mathsf{a}) \geq Q(\mathsf{b}). \tag{8}$$

Consider a BPA m for r, and suppose $s \subset r$. A marginal of m for s is denoted $m^{\downarrow s}$ (defined in Eq. (9)). A similar notation is used for *projections*. For $a \in \Omega_r$, $a^{\downarrow s}$ denote the element of Ω_s that is obtained from a by omitting the values of variables from $r \setminus s$. Similarly, for subset $\mathsf{b} \subseteq \Omega_r$, its projection $\mathsf{b}^{\downarrow s} = \{a^{\downarrow s} : a \in \mathsf{b}\}$. The projection of sets enables us to define a *join* of two sets. Consider two arbitrary sets r and s of variables (they may be disjoint or overlapping, or one may be a subset of the other), and $\mathsf{a} \subseteq \Omega_r$, $\mathsf{b} \subseteq \Omega_s$. Their *join* is defined as:

$$\mathsf{a} \bowtie \mathsf{b} = \{c \in \Omega_{r \cup s} : c^{\downarrow r} \in \mathsf{a} \ \ \& \ \ c^{\downarrow s} \in \mathsf{b}\}.$$

Notice that if r and s are disjoint, then $\mathsf{a} \bowtie \mathsf{b} = \mathsf{a} \times \mathsf{b}$, if $r = s$, then $\mathsf{a} \bowtie \mathsf{b} = \mathsf{a} \cap \mathsf{b}$, and, in general, for $\mathsf{c} \subseteq \Omega_{r \cup s}$, c is a subset of $\mathsf{c}^{\downarrow r} \bowtie \mathsf{c}^{\downarrow s}$, which may be a proper one.

For BPA m for r and $s \subseteq r$, the marginal BPA $m^{\downarrow s}$ for s is defined as follows:

$$m^{\downarrow s}(\mathsf{b}) = \sum_{\mathsf{a} \subseteq \Omega_r :\, \mathsf{a}^{\downarrow s} = \mathsf{b}} m(\mathsf{a}), \tag{9}$$

for all $\mathsf{b} \subseteq \Omega_s$.

An important operator of the Dempster-Shafer (D-S) theory is *Dempster's combination rule*, which combines distinct belief functions. Consider two BPAs m_1 and m_2 for r and s, respectively, and assume they are distinct (independent). Dempster's combination rule is defined for each $\mathsf{c} \subseteq \Omega_{r\cup s}$ as follows:

$$(m_1 \oplus m_2)(\mathsf{c}) = \frac{1}{K} \sum_{\mathsf{a}\subseteq\Omega_r,\, \mathsf{b}\subseteq\Omega_s:\, \mathsf{a}\bowtie\mathsf{b}=\mathsf{c}} m_1(\mathsf{a}) \cdot m_2(\mathsf{b}), \tag{10}$$

where the normalization constant

$$K = \sum_{\mathsf{a}\subseteq\Omega_r,\, \mathsf{b}\subseteq\Omega_s:\, \mathsf{a}\bowtie\mathsf{b}\neq\emptyset} m_1(\mathsf{a}) \cdot m_2(\mathsf{b}). \tag{11}$$

$(1 - K)$ can be interpreted as the *amount of conflict* between m_1 and m_2. If $(1 - K) = 1$, we say that BPAs m_1 and m_2 are in total conflict, and their Dempster's combination is undefined. The assumption of distinct BPAs is fundamental because Dempster's combination is not idempotent[1].

It is known that Dempster's combination is commutative and associative [6]. Another important property of Dempster's combination rule relates to the marginalization of joint BPAs. This property, called *local computation* in [8], says that for m_1 and m_2 defined for r and s, respectively,

$$(m_1 \oplus m_2)^{\downarrow t} = m_1^{\downarrow t} \oplus m_2, \tag{12}$$

if $s \subseteq t \subseteq r \cup s$.

When introducing conditioning for belief functions, Shafer, in his seminal book [6], starts by describing how Dempster's rule of combination makes describing the assimilation of new evidence possible. More than its role for "updating" the evidence, we emphasize in this paper its power to describe knowledge in a form appropriate for belief-function directed graphical models, which generalize probabilistic graphical models called Bayesian networks. This topic is described in the next section.

3 Removal Operator and Conditioning

Belief-function directed graphical models use low-dimensional conditional belief functions (conditionals) as basic building blocks of multidimensional BPAs. Compositional models defined in the next section are composed of (unconditional) low-dimensional BPAs. However, to avoid double-counting of knowledge, we have to compute conditionals from some of them. Therefore, in both these ways of the efficient representation of multidimensional BPAs, we need conditional BPAs.

Consider two BPAs, m_1 for r and m_2 for s. Assume they are marginals of some BPA m, defined for variables $r \cup s$. This means that if r and s are not disjoint, one cannot expect m_1 and m_2 to be distinct. Still, for compositional models, we need to combine them. One cannot use Dempster's rule of combination unless double-counting is prevented. For this reason, we introduce an operator that is an *inverse* to Dempster's rule of combination.

We use the fact that Dempster's combination rule can be expressed in terms of CFs. Let Q_{m_1} and Q_{m_2} be commonality functions of BPAs m_1 and m_2 from Eq. (10). As shown in [6], the CF $Q_{m_1\oplus m_2}$ of their Dempster's combination can be computed for each $\emptyset \neq \mathsf{c} \subseteq \Omega_{r\cup s}$ using the product formula:

$$Q_{m_1\oplus m_2}(\mathsf{c}) = \frac{1}{L} Q_{m_1}(\mathsf{c}^{\downarrow r}) \cdot Q_{m_2}(\mathsf{c}^{\downarrow s}), \tag{13}$$

where the normalization constant

$$L = \sum_{\emptyset\neq\mathsf{c}\subseteq\Omega_{r\cup s}} (-1)^{|\mathsf{c}|+1} Q_{m_1}(\mathsf{c}^{\downarrow r}) \cdot Q_{m_2}(\mathsf{c}^{\downarrow s}) \tag{14}$$

equals the normalization constant K from Eq. (11).

Eq. (13) enables us to define the inverse of Dempster's combination rule called *removal* in [1] (in [9], it is called the *decombination* operator). Consider BPA m for $r \supseteq s$. By removing $m^{\downarrow s}$ from m, we understand the computation of a BPA $\bar{m}$ for r, for which $\bar{m} \oplus m^{\downarrow s} = m$. Since the combination is defined as the pointwise multiplication of CFs followed by normalization, the removal can be defined as the pointwise division of CFs followed by normalization. It means that for the corresponding CF Q_m,

$$(Q_m \ominus Q_{m^{\downarrow s}})(\mathsf{a}) = L^{-1}\, Q_m(\mathsf{a})/Q_{m^{\downarrow s}}(\mathsf{a}^{\downarrow s}) \tag{15}$$

[1]An operator $\oplus$ is said to be idempotent if $m \oplus m = m$ for all m. Nevertheless, $m \oplus m = m$ holds only for some BPAs, e.g., BPAs with several disjoint focal elements, which are all assigned the same value. The idea of distinct belief functions corresponds to no double-counting of non-idempotent knowledge. See a detailed discussion of this notion in [7].

should hold for all nonempty $\mathsf{a} \subseteq \Omega_r$. In this case, the normalization constant L equals

$$L = \sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_r} (-1)^{|\mathsf{a}|+1} Q_m(\mathsf{a}) / Q_{m^{\downarrow s}}(\mathsf{a}^{\downarrow s}). \tag{16}$$

Notice that we want to define the removal only when we remove a marginal $Q_{m^{\downarrow s}}$ from Q_m. Thus, if $Q_{m^{\downarrow s}}(\mathsf{a}^{\downarrow s}) = 0$, then also $Q_m(\mathsf{a}) = 0$. So, Eq. (15) does not uniquely specify the value of $(Q_m \ominus Q_{m^{\downarrow s}})(\mathsf{a})$ for those a, for which in Eq. (15) we get indefinite expression $0/0$. In such situations, we have to assign the value not to violate Eq. (8) expressing the fact that CF $(Q_m \ominus Q_{m^{\downarrow s}})$ should be non-increasing function. It may happen that even this requirement does not specify the value for some states uniquely[2]. In this case, we assign the maximum possible value. The above considerations summarize in the following formal definition.

Definition 1 *Let m be a BPA for r, Q_m denote the corresponding CF, and $s \subset r$. Denote by R an auxiliary function $R : 2^{\Omega_r} \to [0,1]$*

$$R(a) = \begin{cases} Q_m(a)/Q_{m^{\downarrow s}}(a^{\downarrow s}) & \text{if } Q_{m^{\downarrow s}}(a^{\downarrow s}) > 0, \\ \min\left\{Q_m(b)/Q_{m^{\downarrow s}}(b^{\downarrow s}) \ : \ \mathsf{a} \supseteq b \subseteq \Omega_r\right\} \cup \{1\} & \text{otherwise}, \end{cases} \tag{17}$$

and by Q its normalized version $Q = R/L$ (where $L = \sum_{\emptyset \neq \mathsf{a} \subseteq \Omega_r} (-1)^{|\mathsf{a}|+1} R(\mathsf{a})$), which is generally a pseudo-CF. If $Q^{\downarrow s}$ is vacuous, then we call CF Q conditional *CF and denote it $Q_m \ominus Q_m^{\downarrow s} = Q$. If $Q^{\downarrow s}$ is not vacuous, then $Q_m \ominus Q_m^{\downarrow s}$ is undefined.*

Remark Despite the fact that we define the removal operator and the conditional for CFs, in what follows, we also use them for BPAs. Thus, $m^{r|s} = m \ominus m^{\downarrow s}$ denotes the BPA corresponding to CF $Q_m \ominus Q_{m^{\downarrow s}}$. It means that $m \ominus m^{\downarrow s}$ can be computed from $Q_m \ominus Q_{m^{\downarrow s}}$ using Eq. (4). Notice that it can also be computed in another way. Applying Eq. (4) directly to function R, i.e.. computing the function

$$h(\mathsf{a}) = \sum_{\mathsf{b} \subseteq \Omega_r : \, \mathsf{b} \supseteq \mathsf{a}} (-1)^{|\mathsf{b} \setminus \mathsf{a}|} R(\mathsf{b})$$

for all $\mathsf{a} \subseteq \Omega_r$, we obviously get a function $h : 2^{\Omega_r} \to [0,1]$ which, after a possible normalization, equals $m \ominus m^{\downarrow s} = h(\mathsf{a})/L$.

We emphasize that it may happen that $m \ominus m^{\downarrow s}$ has negative masses – then, it is not a BPA and we call it a *pseudo-BPA*. We call $m \ominus m^{\downarrow s}$ a conditional BPA for $r \setminus s$ given s. We do it only when $(m \ominus m^{\downarrow s})^{\downarrow s}$ is vacuous, which is generally the necessary condition for conditionals [10].

Example 1 *Consider variables X and Y with $\Omega_X = \{x, \bar{x}\}$, $\Omega_Y = \{y, \bar{y}\}$. Consider BPA $m_{X,Y}$ for $\{X, Y\}$ with two focal elements: $m_{X,Y}(\{(x,y)\}) = 0.9$ and $m_{X,Y}(\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}) = 0.1$. Its marginal BPA $(m_{X,Y})^{\downarrow X} = m_X$ has also two focal elements: $m_X(\{x\}) = 0.9$, $m_X(\{x,\bar{x}\}) = 0.1$. The computation of $Q_{m_{X,Y}} \ominus Q_{m_X}$ is shown in Table 1. To save room in the heading of the table, we slightly modified the notation: m_X denotes $m_{X,Y}^{\downarrow X}$, and $m_X^{\uparrow\{X,Y\}}$ denotes $m_X \oplus \iota_Y$. Notice also that $Q_{m_{X,Y}} \ominus Q_{m_X} = Q_{m_{X,Y}}/Q_{m_X} = Q_{m_{X,Y}}/Q_{m_X^{\uparrow\{X,Y\}}}$. The last column in the table is the pseudo-BPA corresponding to $Q_{m_{X,Y}} \ominus Q_{m_X}$ computed using Eq. (4).*

Simple Facts about the Removal Operator

Suppose m is a BPA defined for r, and $s \subset r$.

1. $m \ominus m$ is always defined, $m \ominus m$ is vacuous.
2. If $(m \ominus m^{\downarrow s})$ is defined, then $m^{\downarrow s} \oplus (m \ominus m^{\downarrow s}) = m$.
3. If $(m \ominus m^{\downarrow s})^{\downarrow s}$ is a (non-negative) BPA, then for any focal element a of $(m \ominus m^{\downarrow s})$, $\mathsf{a}^{\downarrow s} = \Omega_s$.
4. For a deterministic m, $m^{\downarrow s} \oplus (m^{\downarrow r \setminus s} \oplus \iota_s) = m$.

[2]The reader can easily show it occurs for BPAs, for which Dempster's rule is idempotent ($m \oplus m = m$), which holds for BPAs, the focal element of which are disjoint and all of them are assigned the same value. The examples are deterministic BPAs and uniform Bayesian BPA. In this case, $m \oplus m = m$, as well as $m \oplus \iota_m = m$. Naturally, we do not expect $m \ominus m = m$. We prefer that $m \ominus m = \iota_m$ holds for all m,

Table 1: The computation of $m_{X,Y} \ominus m_X$ in Example 1. Empty cell values equal 0.

$2^{\Omega_{X,Y}}$	$m_{X,Y}$	$m_X^{\uparrow\{X,Y\}}$	$Q_{m_{X,Y}}$	$Q_{m_X^{\uparrow\{X,Y\}}}$	$Q_{m_{X,Y}}/Q_{m_X^{\uparrow\{X,Y\}}}$	$m_{X,Y} \ominus m_X$
$\emptyset$			1	1	1	
$\{(x,y)\}$	0.9		1	1	1	0.9
$\{(x,\bar{y})\}$			0.1	1	0.1	
$\{(\bar{x},y)\}$				0.1		
$\{(\bar{x},\bar{y})\}$			0.1	0.1	1	
$\{(x,y),(x,\bar{y})\}$		0.9	0.1	1	0.1	-0.9
$\{(x,y),(\bar{x},y)\}$				0.1		
$\{(x,y),(\bar{x},\bar{y})\}$			0.1	0.1	1	
$\{(x,\bar{y}),(\bar{x},y)\}$				0.1		
$\{(x,\bar{y}),(\bar{x},\bar{y})\}$			0.1	0.1	1	
$\{(\bar{x},y),(\bar{x},\bar{y})\}$				0.1		
$\{(x,y),(x,\bar{y}),(\bar{x},y)\}$				0.1		
$\{(x,y),(x,\bar{y}),(\bar{x},\bar{y})\}$	0.1		0.1	0.1	1	1
$\{(x,y),(\bar{x},y),(\bar{x},\bar{y})\}$				0.1		
$\{(x,\bar{y}),(\bar{x},y),(\bar{x},\bar{y})\}$				0.1		
$\Omega_{X,Y}$		0.1		0.1		

5. For a deterministic m with a focal element a, the conditional $(m \ominus m^{\downarrow s})$ is a deterministic BPA with a focal element $\mathsf{a}^{\downarrow r\setminus s} \times \Omega_s$.

Proofs: Let Q_m be CF corresponding to BPA m.

Ad. 1. For any $\mathsf{a} \subseteq \Omega_r$, for which $Q_m(\mathsf{a}) > 0$, $R(\mathsf{a}) = 1$, and therefore $R(\mathsf{a}) = 1$ for all $\mathsf{a} \subseteq \Omega_r$, which equals the normalized CF for vacuous BFA.

Ad. 2. For $\emptyset \neq \mathsf{a} \subseteq \Omega_r$ such that $Q_m(\mathsf{a}) > 0$

$$\begin{aligned}\left(Q_m^{\downarrow s} \oplus (Q_m \ominus Q_{m^{\downarrow s}})\right)(\mathsf{a}) &= \frac{1}{K}\left(Q_m^{\downarrow s} \cdot (Q_m \ominus Q_{m^{\downarrow s}})\right)(\mathsf{a}) \\ &= \frac{1}{K}\left(Q_m^{\downarrow s} \cdot \left(\frac{1}{L} \cdot \frac{Q_m}{Q_{m^{\downarrow s}}}\right)\right)(\mathsf{a}) = \frac{1}{K \cdot L} Q_m(\mathsf{a}).\end{aligned}$$

Since if $Q_m(\mathsf{a}) = 0$, then also $\left(Q_m^{\downarrow s} \oplus (Q_m \ominus Q_{m^{\downarrow s}})\right)(\mathsf{a}) = 0$, the product of normalization constants $K \cdot L$ must equal 1 because both Q_m and $\left(Q_m^{\downarrow s} \oplus (Q_m \ominus Q_{m^{\downarrow s}})\right)$ are normalized CFs.

Ad. 3. If there were a focal element $\mathsf{a} \subsetneq \Omega_r$ of $(m \ominus m^{\downarrow s})$, then $\mathsf{a}^{\downarrow s}$ would be a focal element of $(m \ominus m^{\downarrow s})^{\downarrow s}$, which contradicts to $(m \ominus m^{\downarrow s})^{\downarrow s} = \iota_s$.

Ad. 4. Consider a deterministic BPA m with a focal element c, and $\mathsf{a} \subseteq \Omega_s$, $\mathsf{b} \subseteq \Omega_r$. Then

$$m^{\downarrow s}(\mathsf{a}) \cdot (m^{\downarrow r\setminus s} \oplus \iota_s)(\mathsf{b}) = 1$$

only if $\mathsf{a} = \mathsf{c}^{\downarrow s}$ and $\mathsf{b}^{\downarrow r\setminus s} = \mathsf{c}^{\downarrow r\setminus s}$, or, equivalently, if $\mathsf{a} \bowtie \mathsf{b} = \mathsf{c}$. Otherwise, this product equals 0. Therefore, due to Eq. (10), $m^{\downarrow s} \oplus (m^{\downarrow r\setminus s} \oplus \iota_s)$ is a deterministic BPA with focal element c.

Ad. 5. CF Q_m for a deterministic BPA m with a focal element c is

$$Q_m(\mathsf{a}) = \begin{cases} 1 & \text{if } \mathsf{a} \subseteq \mathsf{c}, \\ 0 & \textit{otherwise}, \end{cases} \tag{18}$$

and therefore

$$Q_m(\mathsf{a})/Q_{m^{\downarrow s}}(\mathsf{a}^{\downarrow s}) = \begin{cases} 1 & \textit{if } \mathsf{a} \subseteq \mathsf{c}, \\ 0 & \textit{if } \mathsf{a}^{\downarrow s} \subseteq \mathsf{c}^{\downarrow s} \ \& \ \mathsf{a} \not\subseteq \mathsf{c}, \\ 0/0 & \textit{otherwise}. \end{cases}$$

Using Eq. (17), we get that function R (and therefore also $Q_m \ominus Q_{m^{\downarrow s}}$) equals 0 for $\mathsf{a} \subseteq \Omega_r$, for which $\mathsf{a}^{\downarrow s} \subseteq \mathsf{c}^{\downarrow s}$, and simultaneously $\mathsf{a} \not\subseteq \mathsf{c}$, which occurs when $\mathsf{a}^{\downarrow r\setminus s} \not\subseteq \mathsf{c}^{\downarrow r\setminus s}$. Otherwise, it equals 1 regardless whether $Q_m/Q_{m^{\downarrow s}}$ is equal 1 or it is the indefinite expression $0/0$. Thus, R equals CF of the deterministic BPA with focal element $\mathsf{c}^{\downarrow r\setminus s}$. $\square$

Open Problems

- **Does $(m \ominus m^{\downarrow s})^{\downarrow s} = \iota_s$ hold for all BPAs m?**
  We conjecture that in Definition 1, the assumption that $Q^{\downarrow s}$ is vacuous is unnecessary, that it holds for all BPAs m. If not, for which BPAs this equality holds?

- **Is it possible to compute conditionals without transforming BF into the corresponding CF?**
  In computations, we represent knowledge using BPAs as the list of focal elements, the number of which is usually very small. However, when we convert a BPA to a corresponding CF, the CF is usually non-zero for all subsets of the state space. If we want to compute the conditional using Definition 1, then we have to assess the values of the function R for all states. For example, in Table 1, even though $m_X^{\uparrow\{X,Y\}}$ has only two focal elements, the corresponding CF in column five is non-zero for all subsets of $\Omega_{X,Y}$. This is true because $m_X(\Omega_X) = 0.1 > 0$. So, computations of a conditional using Definition 1 are of high computational complexity and can only be computed for three or four-dimensional BPAs.

  As shown in Simple Facts above, the conditional for a deterministic BPA can be easily obtained. Does a more general class of BPAs exist for which one can compute conditionals directly without enumerating the corresponding CFs?

- **Is it possible to characterize BPAs m, which can be factored as Dempster's combination of its marginal and the corresponding conditional BPA?**
  In probability theory, a joint distribution $P_{X,Y}$ can always be factored into marginal $P_X = (P_{X,Y})^{\downarrow X}$ and a conditional $P_{Y|X}$ such that $P_{X,Y} = P_X \cdot P_{Y|X}$. This is not always true for belief functions. Because of the great computational complexity of the respective algorithms, it would be useful to recognize when such a factorization does not exist. In [11], we proved that $Q_{m_{X,Y}} \ominus Q_{m_X}$ is a CF if and only if there exists a BPA $\hat{m}$ for $\{X, Y\}$ such that $m_{X,Y} = m_X \oplus \hat{m}$, and $\hat{m}^{\downarrow X}$ is the vacuous BPA for X. Nevertheless, the question remains about recognizing whether such $\hat{m}$ exists.

- **Computational problems.**
  As mentioned, when transforming a commonality function to a corresponding BPA, we usually deal with an enormous number of sets. For each of them, we have to find its supersets. The transformation itself (Eq. (4)) is a Möbius transform - i.e., repeated addition and subtraction of many, usually very small numbers. This process often leads to rounding errors. As a rule, it does not happen in the inverse transformation (Eq. (2)) because we handle BPAs with only a few focal elements. So the question is whether there is a class of CFs for which a suitable representation in computer memory would resolve these issues.

4 Compositional Models

Consider two BPAs, m_1 for r and m_2 for s. Assume they are marginals of some BPA m, defined for variables $r \cup s$. Naturally, there is no way how to reconstruct m from its marginals m_1, m_2. However, if we accept the assumption that there is a relation of conditional independence between the considered variables, there may be a unique BPA with the given marginals. In the considered case, it would be the assumption that variables $r \setminus s$ and variables $s \setminus r$ are conditionally independent given variables $r \cap s$. First, we define the notion of conditional independence for BPAs from [1].

Definition 2 *Consider three disjoint sets of variables r, s, t, and a BPA m defined for variables containing $r \cup s \cup t$. Assume r and s are nonempty. We say r and s are conditionally independent given t, with respect to m, written as $r \perp\!\!\!\perp_m s | t$, if there exist BPAs m_1 for $r \cup t$ and m_2 for $s \cup t$ such that $m^{\downarrow r \cup s \cup t} = m_1 \oplus m_2$.*

In the above definition, if t is empty, r and s are said to be unconditionally independent, and the joint BPA $m^{\downarrow r \cup s}$ is equal to Dempster's combination of its marginals. If $t \neq \emptyset$, then one cannot combine the marginal for $r \cup t$ with the marginal for $s \cup t$ because the marginal for variables t would be counted twice – recall that Dempster's combination rule is not idempotent. To avoid double-counting this marginal, one has to use the composition instead of Dempster's rule of combination. For the reasons explained later, we call it a d-composition. It is derived from Dempster's combination rule in [12] and defined as follows.

Definition 3 *Consider BPAs m_1 for r and m_2 for s. Their* d-composition $m_1 \triangleright_d m_2$ *is defined as*

$$m_1 \triangleright_d m_2 = m_1 \oplus (m_2 \ominus m_2^{\downarrow r\cap s}),$$

if the right-hand side of this equality is a BPA. Otherwise $m_1 \triangleright_d m_2$ is undefined.

Notice that this paper excludes the possibility of composing BPAs that would yield a pseudo-BPA (with negative values). Nevertheless, we admit situations when the expression $(m_2 \ominus m_2^{\downarrow r\cap s})$ defining the composition is not a conditional BPA when it is only a pseudo-BPA.

Example 1 (continued) *Dempster's combination of one-dimensional BPA m_X with two focal elements $m_X(\{x\}) = 0.9$, $m_X(\{x, \bar{x}\}) = 0.1$, and pseudo-BPA $(m_{X,Y} \ominus m_{X,Y}^{\downarrow X})$ from the last column of Table 1 yields BPA $m_{X,Y}$ from the first column of Table 1, i.e., for the (pseudo-)BPAs*

$$m_X \oplus (m_{X,Y} \ominus m_{X,Y}^{\downarrow X}) = m_{X,Y}.$$

Similarly, the reader can show that Dempster's combination of pseudo-BPA $(m_{X,Y} \ominus m_{X,Y}^{\downarrow X})$ with any positive one-dimensional Bayesian BPA m_X results in a BPA. Nevertheless, considering other one-dimensional BPAs m_X, their Dempster's combination with pseudo-BPA $(m_{X,Y} \ominus m_{X,Y}^{\downarrow X})$ may yield pseudo-BPAs. For example, $\iota_X \oplus (m_{X,Y} \ominus m_{X,Y}^{\downarrow X}) = (m_{X,Y} \ominus m_{X,Y}^{\downarrow X})$.

In [13], another composition operator for belief functions was introduced. This operator is called the f-composition operator in this paper.

Definition 4 *Consider BPAs m_1 for r and m_2 for s. Their f-composition is a BPA $m_1 \triangleright_f m_2$ defined for each nonempty $\mathsf{c} \subseteq \Omega_{r\cup s}$ by one of the following expressions:*

(i) *if $m_2^{\downarrow r\cap s}(\mathsf{c}^{\downarrow r\cap s}) > 0$ and $\mathsf{c} = \mathsf{c}^{\downarrow r} \bowtie \mathsf{c}^{\downarrow s}$, then* $(m_1 \triangleright_f m_2)(\mathsf{c}) = \dfrac{m_1(\mathsf{c}^{\downarrow r}) \cdot m_2(\mathsf{c}^{\downarrow s})}{m_2^{\downarrow r\cap s}(\mathsf{c}^{\downarrow r\cap s})}$;

(ii) *if $m_2^{\downarrow r\cap s}(\mathsf{c}^{\downarrow r\cap s}) = 0$ and $\mathsf{c} = \mathsf{c}^{\downarrow r} \times \Omega_{s\setminus r}$, then $(m_1 \triangleright_f m_2)(\mathsf{c}) = m_1(\mathsf{c}^{\downarrow r})$;*

(iii) *in all other cases, $(m_1 \triangleright_f m_2)(\mathsf{c}) = 0$.*

An important difference between this definition and the definition of d-composition is visible at first sight. The reader can see that one and only one expression applies for each $\mathsf{c} \subseteq \Omega_{r\cup s}$. Therefore, f-composition is defined for any couple of belief functions. This is its indisputable advantage. A disadvantage is that from the viewpoint of D-S theory, there is no connection to Dempster's rule of combination. The f-composition does not guarantee an expected conditional independence relation among the variables.

However, what is important, both the composition operators introduced in Definitions 3 and 4 satisfy the following properties (properties 1. - 4. are sometimes considered axioms for composition). For proofs, see [13] and [12], where also other properties are studied, including property 5. which also holds for Dempster's combination rule.

Proposition 1 *For both composition operators (d-composition and f-composition) the following statements hold. Assume that BPAs m_r, m_s, and m_t are for r, s, and t, respectively, and that all the d-compositions are defined. Then,*

1. (Domain): *$m_r \triangleright m_s$ is a BPA for variables $r \cup s$.*
2. (Composition preserves first marginal): $(m_r \triangleright m_s)^{\downarrow r} = m_r$.
3. (Commutativity under consistency): *If m_r and m_s are consistent, i.e., $m_r^{\downarrow r\cap s} = m_s^{\downarrow r\cap s}$, then $m_r \triangleright m_s = m_s \triangleright m_r$.*
4. (Associativity under special condition): *If $r \supseteq (s \cap t)$, or, $s \supseteq (r \cap t)$ then, $(m_r \triangleright m_s) \triangleright m_t = m_r \triangleright (m_s \triangleright m_t)$.*
5. (Local computation): *If $(r \cap s) \subseteq t \subseteq (r \cup s)$, then $(m_r \triangleright m_s)^{\downarrow t} = m_r^{\downarrow r\cap t} \triangleright m_s^{\downarrow s\cap t}$.*

Unlike Dempster's rule, which can be applied only to a couple of distinct BPAs, the composition operators are typically used to compose two non-distinct marginals with a non-empty intersection, to assemble two pieces of evidence with some common knowledge. The composition operator is defined to avoid double counting of evidence from the two composed pieces of evidence. Thus, composition and Dempster's combination are designed for different purposes and possess different properties. While Dempster's rule is always commutative and associative, the composition operator meets these properties only in particular situations (see properties 3. and 4. from Proposition 1). On the other hand, Dempster's rule does not preserve the first marginal; it is not idempotent.

Consider BPAs m_1 for r and m_2 for s such that $m_2 \ominus m_2^{\downarrow r\cap s}$ is a BPA. In connection with Definition 3, we will identify situations when conditional BPA $m_2 \ominus m_2^{\downarrow r\cap s}$ is, in a way, "adapted" to BPA m_1. We say that $m_2 \ominus m_2^{\downarrow r\cap s}$ is *tight* with respect to m_1 if for all couples of focal elements a and b (a is a focal element of m_1, and b is a focal element of $m_2 \ominus m_2^{\downarrow r\cap s}$) the following condition holds:

$$\text{for } \forall\, b \in \mathsf{b},\ \exists\, a \in \mathsf{a},\ \text{such that } \{a\} \bowtie \{b\} \neq \emptyset. \tag{19}$$

Expression (ii) in Definition 4 applies to states for which the composed BPAs are, in a way, incompatible; the second argument does not bear the information on how to divide the mass assigned to a focal element of the first argument. Therefore, Expression (ii) assigns the respective value of a mass function to the least specific focal element. The acceptance of this idea makes the f-composition of any couple of BPAs possible. Notice that if the conditional of $m_{Y,Z}$ is tight with respect to $m_{X,Y}$, then Expression (ii) does not find its use.

Facts about the operators of composition (proved in [11])

Suppose m_1 and m_2 are BPAs defined for r and s, respectively.

1. If $m_2 \ominus m_2^{\downarrow r\cap s}$ is a non-negative BPA, then BPA $m_2 \ominus m_2^{\downarrow r\cap s}$ is *tight* with respect to m_1 if and only if $m_1 \triangleright_f m_2 = m_1 \triangleright_d m_2$.

2. If $m_1 \triangleright_d m_2$ is defined, then $Bel_{m_1 \triangleright_f m_2} \leq Bel_{m_1 \triangleright_d m_2}$.

Example 2 *In this example, we present a pair of BPAs, for which the f-composition and d-composition differ. Notice that $m_{Y,Z}$ from Table 2 is a conditional, because $m_{Y,Z}^{\downarrow Y}$ is vacuous, and thus $m_{Y,Z} \ominus (m_{Y,Z})^{\downarrow Y} = m_{Y,Z}$. Notice also that $m_{Y,Z}$ is not tight with respect to $m_{X,Y}$. Their compositions $m_{X,Y} \triangleright_d m_{Y,Z}$ and $m_{X,Y} \triangleright_f m_{Y,Z}$ (see Table 2) differ only in the fact that the d-composition assigns mass 0.70 to $\{(\bar{x}\bar{y}z), (x\bar{y}z)\}$. In contrast, the f-composition assigns this mass to $\{(\bar{x}\bar{y}\bar{z}), (\bar{x}\bar{y}z), (x\bar{y}\bar{z}), (x\bar{y}z)\}$ by Expression (ii). Thus, as more precisely expressed in the assertion above, the result of the f-composition is less specific than that of the d-composition. By the loss of specificity, we have to pay for the ability to combine any couple of BPAs. In other words, when we want to compose two BPAs whose d-composition is undefined, we can do it using f-composition, but we have to reconcile to a partial loss of information.*

Table 2: An example illustrating the relation between $\triangleright_d$ and $\triangleright_f$

a	$m_{X,Y}(a)$
$\{(\bar{x}\bar{y}), (x\bar{y})\}$	0.70
$\{(\bar{x}\bar{y}), (\bar{x}y), (x\bar{y})\}$	0.30

a	$m_{Y,Z}(a)$
$\{(\bar{y}z), (y\bar{z})\}$	0.51
$\{(\bar{y}z), (yz)\}$	0.49

	$(m_{X,Y} \triangleright m_{Y,Z})(a)$	
a	$\triangleright_d$	$\triangleright_f$
$\{(\bar{x}\bar{y}z), (x\bar{y}z)\}$	0.70	
$\{(\bar{x}\bar{y}\bar{z}), (\bar{x}\bar{y}z), (x\bar{y}\bar{z}), (x\bar{y}z)\}$		0.70
$\{(\bar{x}\bar{y}z), (\bar{x}y\bar{z}), (x\bar{y}z)\}$	0.15	0.15
$\{(\bar{x}\bar{y}z), (\bar{x}yz), (x\bar{y}z)\}$	0.15	0.15

Open Problems

- **What are the necessary and sufficient conditions for $m_1 \triangleright_d m_2 = m_1 \triangleright_f m_2$?**
  Fact 1 characterize situations when $m_1 \triangleright_d m_2 = m_1 \triangleright_f m_2$ under the assumption that $m_2 \ominus m_2^{\downarrow r \cap s}$ is a non-negative BPA. How is it for situations when $m_2 \ominus m_2^{\downarrow r \cap s}$ is a pseudo-PBA?

- **Is it possible to characterize pairs of BPAs m_1 and m_2 for which the d-composition yields a non-negative BPA?**
  Consider BPAs m_1 for r and m_2 for s. If $m_2 \ominus m_2^{\downarrow r \cap s}$ is a BPA, then $m_1 \triangleright_d m_2$ is also a BPA. This is a sufficient condition. But, as shown in Example 1, it is not necessary.

- **Given BPA m for $r \supsetneq s$, and its conditional pseudo-BPA $m \ominus m^{\downarrow s}$. What is the class of BPAs $\bar{m}$, for which $\bar{m} \triangleright_d m$ is a non-negative BPA.**
  This is a sub-problem of the problem above. Consider, for example $m_{X,Y}$ from Example 1. It is not difficult to show that $\bar{m}_X \triangleright_d m_{XY}$ is a non-negative BPA for any Bayesian $\bar{m}$ for X.

- **Computational problems.**
  Because of the computational problems mentioned in the preceding section, we can currently compute $m_1 \triangleright_d m_2$ only when the dimension of m_2 is not greater than four. For higher dimensions, we have to approximate $m_1 \triangleright_d m_2$ by $m_1 \triangleright_f m_2$. Are there chances to find the representation of conditionals $m_2 \ominus m_2^{\downarrow r \cap s}$ in computer memory so that the computation of $m_1 \triangleright_d m_2$ would be tractable for higher dimensions?

5 Summary

Both graphical and compositional models for belief functions are based on the idea that a multidimensional model is constructed from low-dimensional belief functions, introducing conditional independence relations among the variables. Both models must employ conditional belief functions to avoid double-counting of knowledge. In this paper, we study the possibility of obtaining conditional BPAs by applying the removal operator, an operator that is inverse to Dempster's combination operator. It is associated with two types of problems, some of which remain open. Theoretical problems arise from the fact that, in some situations, the removal operator is undefined. Also, what causes even more severe problems, the result may go beyond the classical belief function theory because the corresponding BPAs may have negative masses. The other problem is connected with the super-exponential computational complexity of the removal operator. Namely, we only know one way to implement the removal operator based on the transformation of BPAs into commonality functions, which becomes intractable if the dimension of the domains of the BPAs exceeds four.

Acknowledgement This research was financially supported by the Czech Science Foundation Grant No. 21-07494S to the first two authors and by the Ronald G. Harper Professorship at the University of Kansas to the third author.

References

[1] P. P. Shenoy. Conditional independence in valuation-based systems. *International Journal of Approximate Reasoning*, 10(3):203–234, 1994.

[2] Inés Couso, Serafín Moral, and Peter Walley. A survey of concepts of independence for imprecise probabilities. *Risk, Decision and Policy*, 5(2):165–181, 2000.

[3] Boutheina Ben Yaghlane, Philippe Smets, and Khaled Mellouli. Belief function independence: I. and II. the marginal and conditional case. *International Journal of Approximate Reasoning*, 29, 31:47–70,31–75, 2002.

[4] Jirina Vejnarová. On conditional independence in evidence theory. In *Proceedings of ISIPTA*, volume 9, pages 431–440. Citeseer, 2009.

[5] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. On the relationship between graphical and compositional models for the Dempster-Shafer theory of belief functions. In E. Miranda, I. Montes, E. Quaeghebeur, and B. Vantaggi, editors, *Proceedings of the 13th International Symposium on Imprecise Probability: Theories and Applications*, volume 215 of *Proceedings of Machine Learning Research*, pages 259–269. MLR Press, 2023.

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

[7] P. P. Shenoy. On distinct belief functions in the Dempster-Shafer theory. In E. Miranda, I. Montes, E. Quaeghebeur, and B. Vantaggi, editors, *Proceedings of the 13th International Symposium on Imprecise Probability: Theories and Applications*, volume 215 of *Proceedings of Machine Learning Research*, pages 426–437. MLR Press, 2023.

[8] P. P. Shenoy and G. Shafer. Axioms for probability and belief-function propagation. In R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanal, editors, *Uncertainty in Artificial Intelligence 4*, volume 9 of *Machine Intelligence and Pattern Recognition Series*, pages 169–198. North-Holland, Amsterdam, Netherlands, 1990.

[9] P. Smets. The canonical decomposition of a weighted belief. In *Proceedings of the 1995 IJCAI Conference*, volume 95, pages 1896–1901, 1995.

[10] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. On conditional belief functions in directed graphical models in the Dempster-Shafer theory. *International Journal of Approximate Reasoning*, 160(9):in press, 2023.

[11] R. Jiroušek, V. Kratochvíl, and P. P. Shenoy. Entropy for evaluation of Dempster-Shafer belief function models. *International Journal of Approximate Reasoning*, 151(12), 2022.

[12] R. Jiroušek and P. P. Shenoy. Compositional models in valuation-based systems. *International Journal of Approximate Reasoning*, 55(1):277–293, 2014.

[13] R. Jiroušek, J. Vejnarová, and M. Daniel. Compositional models for belief functions. In *Proceedings of ISIPTA*, volume 7, pages 243–252, 2007.