On Two Composition Operators in Dempster-Shafer Theory

Radim Jiroušek
Faculty of Management, Univ. of Economics
Jindřichuv Hradec, Czech Republic
radim@utia.cas.cz

Abstract

Efficient computations with probabilistic multidimensional models are made possible if the respective probability measure (distribution) is in the form of a decomposable model. Some of the advantageous properties of these models are based on the fact that factorization and conditional independence coincide. It means that a decomposable multidimensional model can be assembled (composed) from its low-dimensional marginals with the help of an *operator of composition*, which introduces conditional independence relations among the variables.

The problem arises when we also want to apply these ideas in Dempster-Shafer theory of evidence, because two different operators of composition have been introduced in literature. The present paper serves as a survey of results on these two operators, recollects their common properties and differences, and tries to find a proper role for each of them.

Keywords. Factorization, conditional independence, combination, composition, decomposable model, IPFP.

1 Introduction

In every textbook dealing with Bayesian network theory there inevitably appears a basic theorem saying that in probability theory, factorization and conditional independence coincide. To express this property more exactly (and simultaneously in its simplest version), consider a probability measure π defined on a finite three-dimensional Cartesian product space $\mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$. Then

$$\pi(\mathbf{a}) \cdot \pi^{\downarrow\{Z\}}(\mathbf{a}^{\downarrow\{Z\}}) = \pi^{\downarrow\{X,Z\}}(\mathbf{a}^{\downarrow\{X,Z\}}) \cdot \pi^{\downarrow\{Y,Z\}}(\mathbf{a}^{\downarrow\{Y,Z\}}) \quad (1)$$

for all $\mathbf{a} \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$, if and only if there exist two functions

$$\phi : \mathbb{X} \times \mathbb{Z} \longrightarrow \mathbb{R}^+,$$
$$\psi : \mathbb{Y} \times \mathbb{Z} \longrightarrow \mathbb{R}^+,$$

such that

$$\pi(\mathbf{a}) = \phi(\mathbf{a}^{\downarrow\{X,Z\}}) \cdot \psi(\mathbf{a}^{\downarrow\{Y,Z\}}) \quad (2)$$

for all $\mathbf{a} \in \mathbb{X} \times \mathbb{Y} \times \mathbb{Z}$.

The equality (1) says that for probability measure π variables X and Y are conditionally independent given variable Z, and equality (2) expresses the fact that measure π factorizes with respect to cover $\{\mathbb{X} \times \mathbb{Z}, \mathbb{Y} \times \mathbb{Z}\}$.

The importance of the factorization stems from the fact that it describes formal conditions under which it is possible to represent a multidimensional Bayesian network with a reasonable number of parameters (conditional probabilities) and to design computationally tractable inference procedures. On the other hand, the concept of conditional independence is comprehensible to users. Thus, verification of the formal conditions for factorization is made possible by the very fact that these two concepts coincide in probability theory. Namely, expressing a probability distribution in a factorized form is equivalent to introducing a conditional independence relation among the variables. And to verify the model, the users should consider whether the introduced conditional independence relations are justifiable (or at least acceptable).

Trying to uncover a similar relationship between factorization and conditional independence in Dempster-Shafer theory of evidence, one easily comes to the conclusion that conditional independence coincides with the factorization of commonality functions, which are, unfortunately, completely illegible to users. This is one of the reasons why we will focus on factorization of basic probability assignments in this paper. We will show that the notions of conditional independence and factorization (of basic probability assignments) correspond to two composition operators studied previously, in ISIPTA paper [8].

The present paper, which is in fact a synthesis of known results about the two operators of composition in D-S theory, is organized as follows. In the next section the necessary notions and notation are introduced. Section 3 is devoted to the properties of the composition operators and uncovers their pros and cons from the point of view of their computational complexity. In Section 4 we describe (using a simple example) how to make some computations with belief functions tractable. The basic idea is the same as the one used for computations with Bayesian networks. We will show how to represent multidimensional belief functions in the form of decomposable models, for which we will employ both the studied operators of composition. Section 5 explains the role of one of the operators in computation of conditionals.

2 Basic Notions and Notation

In this paper we consider a finite set of finite valued variables $N = \{X_1, X_2, \ldots, X_n\}$; $\mathbb{X}_i$ denotes the set of states of variable X_i. $\mathbb{X}_N = \mathbb{X}_1 \times \mathbb{X}_2 \times \ldots \times \mathbb{X}_n$ denotes a finite multidimensional space of states of variables N, and its subspaces[1] (for all $K \subseteq N$) are denoted by

$$\mathbb{X}_K = \times_{i \in K} \mathbb{X}_i.$$

For a state $x = (x_1, x_2, \ldots, x_n) \in \mathbb{X}_N$ its projection into subspace $\mathbb{X}_K$ is denoted by $x^{\downarrow K} = (x_{i, i\in K})$, and for $\mathbf{a} \subseteq \mathbb{X}_N$

$$\mathbf{a}^{\downarrow K} = \{y \in \mathbb{X}_K : \exists x \in \mathbf{a}, x^{\downarrow K} = y\}.$$

Symbol $2^{\mathbf{a}}$ denotes the set of all nonempty subsets of $\mathbf{a}$. By a *join* of two sets $\mathbf{a} \subseteq \mathbb{X}_K$ and $\mathbf{b} \subseteq \mathbb{X}_L$ we understand a set

$$\mathbf{a} \bowtie \mathbf{b} = \{x \in \mathbb{X}_{K\cup L} : x^{\downarrow K} \in \mathbf{a} \ \ \& \ \ x^{\downarrow L} \in \mathbf{b}\}.$$

Realize that if K and L are disjoint, then $\mathbf{a} \bowtie \mathbf{b} = \mathbf{a} \times \mathbf{b}$, if $K = L$ then $\mathbf{a} \bowtie \mathbf{b} = \mathbf{a} \cap \mathbf{b}$, and, generally, for $\mathbf{c} \subseteq \mathbb{X}_{K\cup L}$, $\mathbf{c}$ is a subset of $\mathbf{c}^{\downarrow K} \bowtie \mathbf{c}^{\downarrow L}$, which may be proper. This is why we will often use the symbol

$$2^{\mathbb{X}_{K\bowtie L}} = \{\mathbf{c} \subseteq \mathbb{X}_{K\cup L} : \mathbf{c} \neq \emptyset \ \& \ \mathbf{c} = \mathbf{c}^{\downarrow K} \bowtie \mathbf{c}^{\downarrow L}\}.$$

Let us mention that sets $\mathbf{c} \in 2^{\mathbb{X}_{K\bowtie L}}$ are called *Z-layered rectangles* in [2].

In what follows it will be important to realize that cardinality of $2^{\mathbb{X}_{K\bowtie L}}$, though growing exponentially with $|\mathbb{X}_{K\cup L}|$, is much smaller than $|2^{\mathbb{X}_{K\cup L}}|$. For example, for binary variables

$$|\mathbb{X}_{\{1,2,3\}}| = 8,$$
$$|2^{\mathbb{X}_{\{1,2,3\}}}| = 255,$$
$$|2^{\mathbb{X}_{\{1,2\}\bowtie\{2,3\}}}| = 99,$$

and for ternary variables

$$|\mathbb{X}_{\{1,2,3\}}| = 27,$$
$$|2^{\mathbb{X}_{\{1,2,3\}}}| = 134\,217\,727,$$
$$|2^{\mathbb{X}_{\{1,2\}\bowtie\{2,3\}}}| = 124\,999.$$

2.1 Belief Functions

The role played in probability theory by probability measures (or probability distributions), is played by belief functions in Dempster-Shafer theory. It is well known [14] that these functions can be equivalently represented in several ways. In this paper we will use just basic probability assignments, and commonality functions.

Basic probability assignment (bpa) on $\mathbb{X}_K$ is a function

$$\mu : 2^{\mathbb{X}_K} \longrightarrow \mathbf{R}.$$

Though most authors require this function to be non-negative and normalized, in this paper we accept the more general approach of Shenoy [15], who does not restrict the class of considered functions and says that bpa is *proper* if this function is non-negative, and further says that it is *normal* if

$$\sum_{\mathbf{a} \subseteq \mathbb{X}_K} \mu(\mathbf{a}) = 1.$$

Nevertheless, not even in this paper will we consider all possible functions. When speaking about a bpa we will always assume that its corresponding *commonality function* (comf), which is a function on $\mathbb{X}_K$ defined for each nonempty $\mathbf{a} \subseteq \mathbb{X}_K$

$$\theta(\mathbf{a}) = \sum_{\mathbf{b} \supseteq \mathbf{a}} \mu(\mathbf{b}), \tag{3}$$

is strictly positive. Recall that this transformation of bpas into comfs is unique, and that a bpa can be reconstructed from its comf using the following formula (Möbius transform – see [14])

$$\mu(\mathbf{a}) = \sum_{\mathbf{b} \supseteq \mathbf{a}} (-1)^{|\mathbf{b} \setminus \mathbf{a}|} \theta(\mathbf{b}), \tag{4}$$

for all nonempty $\mathbf{a} \subseteq \mathbb{X}_K$. This enables us to call comf θ *proper* (*normal*) if the corresponding bpa is proper (normal).

$\mathbf{a} \in 2^{\mathbb{X}_K}$ is said to be a *focal element* of bpa μ if $\mu(\mathbf{a}) \neq 0$. Quite often, the list of focal elements and

[1] In our examples we will use a simplified notation. Instead of a correct notation for a subset of variables, say, $K = \{X_1, X_3, X_7\}$ we will use just $K = \{1, 3, 7\}$.

the respective values of bpa are used for belief function representation. However, as we will see later, it does not mean that a belief function may be represented by the list of values of the respective comf on the focal elements, because comfs are quite often positive also for non-focal elements. The exceptions are so-called *probabilistic* (some authors call them *Bayesian*) belief functions, for which all focal elements are *singletons* (the cardinality of each focal element is one), and, as it can be immediately seen from Formula (3), $\mu = \theta$.

The last notion introduced in this section is that of dominance. Consider two bpas μ_1 and μ_2 defined on the same $\mathbb{X}_K$. We say that μ_1 dominates μ_2 (and write $\mu_1 \gg \mu_2$) if all focal elements of μ_2 are also focal elements of μ_1, i.e., if

$$\mu_1(\mathbf{a}) = 0 \implies \mu_2(\mathbf{a}) = 0$$

for all $\mathbf{a} \in 2^{\mathbb{X}_K}$.

2.2 Combination

An important notion in D-S theory is the notion of combination $\oplus$, usually called *Dempster's rule of combination*.

Definition 1 *Consider two arbitrary bpas μ_1 on $\mathbb{X}_K$ and μ_2 on $\mathbb{X}_L$ ($K \neq \emptyset \neq L$). A* combination $\mu_1 \oplus \mu_2$ *is defined as follows:*
if $\mathbf{c} \in 2^{\mathbb{X}_{K \bowtie L}}$ then

$$(\mu_1 \oplus \mu_2)(\mathbf{c}) = \alpha^{-1} \sum_{\mathbf{a} \subseteq \mathbb{X}_K, \mathbf{b} \subseteq \mathbb{X}_L : \mathbf{a} \bowtie \mathbf{b} = \mathbf{c}} \mu_1(\mathbf{a}) \cdot \mu_2(\mathbf{b})$$

where α is a normalization constant

$$\alpha = \sum_{\mathbf{d} \in 2^{\mathbb{X}_{K \bowtie L}}} \sum_{\mathbf{a} \subseteq \mathbb{X}_K, \mathbf{b} \subseteq \mathbb{X}_L : \mathbf{a} \bowtie \mathbf{b} = \mathbf{d}} \mu_1(\mathbf{a}) \cdot \mu_2(\mathbf{b}),$$

and $(\mu_1 \oplus \mu_2)(\mathbf{c}) = 0$ for all $\mathbf{c} \in 2^{\mathbb{X}_{K \cup L}} \setminus 2^{\mathbb{X}_{K \bowtie L}}$.

It is well known (e.g., [14]) that Dempster's rule of combination can also be (even more elegantly) defined with the help of the respective comfs. Namely, if θ_1, θ_2 correspond to μ_1, μ_2, respectively, then the comf corresponding to $\mu_1 \oplus \mu_2$ is

$$(\theta_1 \oplus \theta_2)(\mathbf{c}) = \frac{\theta_1(\mathbf{c}^{\downarrow K}) \cdot \theta_2(\mathbf{c}^{\downarrow L})}{\sum_{\mathbf{d} \in 2^{\mathbb{X}_{K \cup L}}} (-1)^{|\mathbf{d}|+1} \, \theta_1(\mathbf{d}^{\downarrow K}) \cdot \theta_2(\mathbf{d}^{\downarrow L})}.$$

The introduced operator of combination models a belief update. So, it is not surprising that this operator is not idempotent, which means that, generally, $\mu \oplus \mu \neq \mu$. Indeed, when hearing the same piece of information from two independent sources we get more convinced about its currency.

The reader can easily verify the non-idempotence of Dempster's rule of combination on a simple bpa μ with two focal elements: $\mu(\mathbf{a}) = \frac{1}{3}, \mu(\mathbf{b}) = \frac{2}{3}$. Namely, the resulting bpa $\mu \oplus \mu$ has, naturally, the same two focal elements

$$(\mu \oplus \mu)(\mathbf{a}) = \frac{1}{5}, \qquad (\mu \oplus \mu)(\mathbf{b}) = \frac{4}{5}.$$

Contrary to the belief update, when assembling a global knowledge from its local pieces we need a tool that *is* idempotent. If the local pieces of knowledge are consistent, each of them should be preserved unchanged in its global representation. Using a mathematical terminology, we can also say that in this case we are looking for a *join extension* of the local pieces. And this is the goal for which the operator of composition introduced in the following section is designed.

2.3 Operators of Composition

In this paragraph we introduce two composition operators. Definition 2 is based on Dempster's rule of combination and its equivalence to (normalized) multiplication of the respective comfs. As we showed above, this operation is not idempotent; we thus have to avoid double counting of contributions on the overlapping subspace. In other words, we have to ensure that each piece of local information is considered only once (for details see [15] or [9]). Such a removal of information is performed by an operation inverse to Dempster's rule of combination (division of the respective comfs). Notice that Definition 2 is described with the help of comfs while Definition 3 makes use of bpas. Nonetheless, because of the one-to-one correspondence between bpas and comfs both these operators are naturally extended to both bpas and comfs. Note that Definition 2 is from [9], and Definition 3 first appeared in [10].

Definition 2 *Consider two arbitrary bpas, μ_1 on $\mathbb{X}_K$ and μ_2 on $\mathbb{X}_L$ ($K \neq \emptyset \neq L$) and assume that $\mu_2^{\downarrow K \cap L} \gg \mu_1^{\downarrow K \cap L}$. Let θ_1 and θ_2 be the respective comfs. A* composition $\theta_1 \,{}^{d}\!\triangleright\, \theta_2$ *is defined for each nonempty $\mathbf{c} \subseteq \mathbb{X}_{K \cup L}$ by the following formula:*

$$(\theta_1 \,{}^{d}\!\triangleright\, \theta_2)(\mathbf{c}) = \begin{cases} \alpha^{-1} \frac{\theta_1(\mathbf{c}^{\downarrow K}) \cdot \theta_2(\mathbf{c}^{\downarrow L})}{\theta_2^{\downarrow K \cap L}(\mathbf{c}^{\downarrow K \cap L})} & \textit{if } \theta_2^{\downarrow K \cap L}(\mathbf{c}^{\downarrow K \cap L}) > 0, \\ 0 & \textit{otherwise,} \end{cases}$$

where α is a normalization constant defined as

$$\alpha = \sum_{\mathbf{d} \in 2^{\mathbb{X}_{K \cup L}} : \theta_2^{\downarrow K \cap L}(\mathbf{d}^{\downarrow K \cap L}) > 0} (-1)^{|\mathbf{d}|+1} \frac{\theta_1(\mathbf{d}^{\downarrow K}) \cdot \theta_2(\mathbf{d}^{\downarrow L})}{\theta_2^{\downarrow K \cap L}(\mathbf{d}^{\downarrow K \cap L})}.$$

Definition 3 *Consider two normal bpas, μ_1 on $\mathbb{X}_K$ and μ_2 on $\mathbb{X}_L$ ($K \neq \emptyset \neq L$). A* composition $\mu_1 \overset{f}{\triangleright} \mu_2$ *is defined for each nonempty $\mathbf{c} \subseteq \mathbb{X}_{K\cup L}$ by one of the following expressions:*

(i) *if $\mu_2^{\downarrow K\cap L}(\mathbf{c}^{\downarrow K\cap L}) > 0$ and $\mathbf{c} \in 2^{\mathbb{X}_{K\bowtie L}}$ then*

$$(\mu_1 \overset{f}{\triangleright} \mu_2)(\mathbf{c}) = \frac{\mu_1(\mathbf{c}^{\downarrow K}) \cdot \mu_2(\mathbf{c}^{\downarrow L})}{\mu_2^{\downarrow K\cap L}(\mathbf{c}^{\downarrow K\cap L})};$$

(ii) *if $\mu_2^{\downarrow K\cap L}(\mathbf{c}^{\downarrow K\cap L}) = 0$ and $\mathbf{c} = \mathbf{c}^{\downarrow K} \times \mathbb{X}_{L\setminus K}$ then*

$$(\mu_1 \overset{f}{\triangleright} \mu_2)(\mathbf{c}) = m_1(\mathbf{c}^{\downarrow K});$$

(iii) *in all other cases, $(\mu_1 \overset{f}{\triangleright} \mu_2)(\mathbf{c}) = 0$.*

Having two operators of composition, quite a natural question arises: why do we need both of them? The operators differ in many aspects. As we will discuss in the next section, each of them has different computational complexity and different semantics. Even so, the answer to the previous question is not straightforward; in fact, information found throughout this entire paper should help readers form their own opinions.

To simplify the following exposition, let us make two conventions. First, whenever we use symbol $\triangleright$ then the respective assertion holds for both the defined operators $\overset{d}{\triangleright}$ and $\overset{f}{\triangleright}$. Second, to avoid frequent repetition of the condition on dominance of arguments required in the definition of $\overset{d}{\triangleright}$, and the requirement of normality needed in the definition of $\overset{f}{\triangleright}$, whenever the operators are used in the following text, we will assume that the necessary conditions under which the respective operator is defined are fulfilled.

3 Properties of Operators of Composition

Both the operators of composition introduced in the preceding section comply with the properties expected from composition. These properties were originally proven for probability theory in [6], and later also for Shenoy's Valuation-Based systems (VBS) in [9], from which it follows that all of them hold for D-S theory.

Theorem 1 *Suppose μ_1, μ_2 and μ_3 are bpas on $\mathbb{X}_K$, $\mathbb{X}_L$, and $\mathbb{X}_M$, respectively. Then the following statements hold:*

1. (Domain): *$\mu_1 \triangleright \mu_2$ is a bpa on $\mathbb{X}_{K\cup L}$.*
2. (Composition preserves first marginal): *$(\mu_1 \triangleright \mu_2)^{\downarrow K} = \mu_1$.*
3. (Reduction:) *If $L \subseteq K$ then, $\mu_1 \triangleright \mu_2 = \mu_1$.*
4. (Non-commutativity): *In general, $\mu_1 \triangleright \mu_2 \neq \mu_2 \triangleright \mu_1$.*
5. (Commutativity under consistency): *If μ_1 and μ_2 have a common marginal on $\mathbb{X}_{K\cap L}$, i.e., $\mu_1^{\downarrow K\cap L} = \mu_2^{\downarrow K\cap L}$, then $\mu_1 \triangleright \mu_2 = \mu_2 \triangleright \mu_1$.*
6. (Non-associativity): *In general, $(\mu_1 \triangleright \mu_2) \triangleright \mu_3 \neq \mu_1 \triangleright (\mu_2 \triangleright \mu_3)$.*
7. (Associativity under special condition I): *If $K \supset (L \cap M)$ then, $(\mu_1 \triangleright \mu_2) \triangleright \mu_3 = \mu_1 \triangleright (\mu_2 \triangleright \mu_3)$.*
8. (Associativity under special condition II): *If $L \supset (K \cap M)$ then, $(\mu_1 \triangleright \mu_2) \triangleright \mu_3 = \mu_1 \triangleright (\mu_2 \triangleright \mu_3)$.*
9. (Stepwise combination): *If $(K \cap L) \subseteq M \subseteq L$ then, $(\mu_1 \oplus \mu_2^{\downarrow M}) \triangleright \mu_2 = \mu_1 \oplus \mu_2$.*
10. (Stepwise composition): *If $(K \cap L) \subseteq M \subseteq L$ then, $(\mu_1 \triangleright \mu_2^{\downarrow M}) \triangleright \mu_2 = \mu_1 \triangleright \mu_2$.*
11. (Exchangeability): *If $K \supset (L \cap M)$ then, $(\mu_1 \triangleright \mu_2) \triangleright \mu_3 = (\mu_1 \triangleright \mu_3) \triangleright \mu_2$.*
12. (Simple marginalization): *If $(K\cap L) \subseteq M \subseteq K\cup L$ then, $(\mu_1 \triangleright \mu_2)^{\downarrow M} = \mu_1^{\downarrow K\cap M} \triangleright \mu_2^{\downarrow L\cap M}$.*
13. (Irrelevant combination): *If $M \subseteq K \setminus L$ then, $\mu_1 \triangleright (\mu_2 \oplus \mu_3) = \mu_1 \triangleright \mu_2$.*

From the formal point of view, the main difference between $\overset{d}{\triangleright}$ and $\overset{f}{\triangleright}$ is in their computational complexity. We need not subject them to a detailed and precise complexity analysis, because the difference is visible at the first sight.

3.1 Computational Complexity

Let us start with a simpler task. Consider the formulae in Definition 3, which give direct instructions for how to compute the composition $\overset{f}{\triangleright}$. To compute $\mu_1 \overset{f}{\triangleright} \mu_2$, it is necessary to find out all of its focal elements $\mathbf{c} \in 2^{\mathbb{X}_{K\cup L}}$ and compute the respective value $(\mu_1 \overset{f}{\triangleright} \mu_2)(\mathbf{c})$. Therefore, the computational complexity of this process is linear in the number of those $\mathbf{c} \in 2^{\mathbb{X}_{K\cup L}}$ that must be checked to determine whether they are focal elements. There are two direct ways in which to do it. One is based on the fact that all focal elements of $\mu_1 \overset{f}{\triangleright} \mu_2$ are from $2^{\mathbb{X}_{K\bowtie L}}$. So, we can generate all the elements of $2^{\mathbb{X}_{K\bowtie L}}$, which is, as we said (and illustrated) in Section 2, substantially smaller than $2^{\mathbb{X}_{K\cup L}}$.

The other, quite often more efficient, possibility is based on the fact that for all focal elements $\mathbf{c} \in 2^{\mathbb{X}_{K\cup L}}$ of $\mu_1 \overset{f}{\triangleright} \mu_2$ the following two conditions must hold simultaneously

(1) $\mu_1(\mathbf{c}^{\downarrow K}) \neq 0$;

(2) either $\mu_2(\mathbf{c}^{\downarrow L}) \neq 0$, or $\mathbf{c}^{\downarrow L} = \mathbf{c}^{\downarrow K\cap L} \times \mathbb{X}_{L\setminus K}$.

This means that the number of potential focal elements of $\mu_1 \overset{f}{\triangleright} \mu_2$ cannot exceed the product of two numbers: the number of focal elements of μ_1 and the number of focal elements of μ_2. In the case when the considered belief functions are represented by the lists of their focal elements, these numbers are usually limited.

For operator $\overset{d}{\triangleright}$ the situation is different. Regardless of the fact that the number of focal elements is also limited, namely, by $|2^{\mathbb{X}_{K\bowtie L}}|$, one can hardly expect that it would be possible to find an algorithm whose computational complexity would be linear in the number of (potential) focal elements. This pessimistic statement holds for Dempster's rule of combination (each value of $\mu_1 \oplus \mu_2$ is computed as a summation), the more it holds for $\mu_1 \overset{d}{\triangleright} \mu_2$. To the best of our knowledge, up to now, no other way to compute $\mu_1 \overset{d}{\triangleright} \mu_2$ has been known than to convert the composed bpas into comfs, and afterward carry out the computations described in Definition 2. However, even for bpa μ_1 with a limited number of focal elements, the number of those $\mathbf{c} \in 2^{\mathbb{X}_K}$ for which $\theta_1(\mathbf{c}) > 0$ holds may be very high. Let us illustrate the situation on an extreme (but appearing in practice) situation.

3.2 Example

Consider bpas μ_1 and μ_2 on $\mathbb{X}_K$, $\mathbb{X}_L$, respectively, and assume they define lower probabilities on the respective subspaces, i.e., their focal elements are only singletons plus the whole $\mathbb{X}_K$ for μ_1, and singletons plus the whole $\mathbb{X}_L$ for μ_2. It means that μ_1 has no more than $|\mathbb{X}_K| + 1$ focal elements, and μ_2 has no more than $|\mathbb{X}_L| + 1$ focal elements.

However, one can immediately see from Formula (3) that any of the corresponding comfs θ_1, θ_2 may easily be positive for all the elements of $2^{\mathbb{X}_K}$, $2^{\mathbb{X}_L}$, respectively. It means that when computing $\theta_1 \overset{d}{\triangleright} \theta_2$ we have to compute the respective values for all $\mathbf{c} \in 2^{\mathbb{X}_{K\cup L}}$.

Computation of $\mu_1 \overset{f}{\triangleright} \mu_2$ is different in this example. If $\mu_2^{\downarrow K\cap L} \gg \mu_1^{\downarrow K\cap L}$ then the resulting bpa $\mu_1 \overset{f}{\triangleright} \mu_2$ has the same property as μ_1 and μ_2: its focal elements are only singletons plus the whole $\mathbb{X}_{K\cup L}$ (this is true because for the considered bpas case **(ii)** of Definition 3 can never assign a value different from 0). If $\mu_2^{\downarrow K\cap L} \not\gg \mu_1^{\downarrow K\cap L}$, i.e., for some $\mathbf{b} \in \mathbb{X}_{K\cap L}$, $\mu_2^{\downarrow K\cap L}(\mathbf{b}) = 0$ and $\mu_1^{\downarrow K\cap L}(\mathbf{b}) \neq 0$, then the number of focal elements $\mathbf{c} \in \mathbb{X}_{K\cup L}$, for which $\mathbf{c}^{\downarrow K\cap L} = \mathbf{b}$ does not exceed

$$|\{\mathbf{a} \in \mathbb{X}_K : \mathbf{a}^{\downarrow K\cap L} = \mathbf{b}\}|,$$

and therefore we see that the number of focal elements of $\mu_1 \overset{f}{\triangleright} \mu_2$ cannot exceed $|\mathbb{X}_{K\cup L}| + 1$.

3.3 Factorization

As a direct consequence of Properties 2 and 5 in Theorem 1, one can see that if two bpas μ_1 and μ_2 (assume again they are defined on $\mathbb{X}_K$ and $\mathbb{X}_L$, respectively) have a common marginal (i.e., $\mu_1^{\downarrow K\cap L} = \mu_2^{\downarrow K\cap L}$), then both $\mu_1 \overset{d}{\triangleright} \mu_2$ and $\mu_1 \overset{f}{\triangleright} \mu_2$ are common extensions of μ_1 and μ_2. Generally, these two extensions differ from each other: each of them has its own semantics.

From the point of view of this paper it is important to say that $\overset{d}{\triangleright}$ reflects the notion of conditional independence in the sense used by Shafer [14] and Shenoy [15]. More precisely, $\mu^{\downarrow K\cup L} = \mu^{\downarrow K} \overset{d}{\triangleright} \mu_2^{\downarrow L}$ holds if and only if variables $K \setminus L$ and $L \setminus K$ are for the considered belief function (bpa μ) *conditionally independent given variables* $K \cap L$.

The semantics of the fact that $\mu^{\downarrow K\cup L} = \mu^{\downarrow K} \overset{f}{\triangleright} \mu_2^{\downarrow L}$ is different. Namely, it is a direct consequence of Lemma 3 in [16] that $\mu^{\downarrow K\cup L} = \mu^{\downarrow K} \overset{f}{\triangleright} \mu^{\downarrow L}$ if and only if $\mu^{\downarrow K\cup L}$ factorizes with respect to a couple $\{K, L\}$ in the sense of the following definition.

Definition 4 *Consider bpa μ on $\mathbb{X}_M$, and two subsets of variables $K, L \subset M$. We say that μ* factorizes *with respect to $\{K, L\}$ if*

(a) $\mu^{\downarrow K\cup L}(\mathbf{c}) = 0$ for all $\mathbf{c} \in (2^{\mathbb{X}_{K\cup L}} \setminus 2^{\mathbb{X}_{K\bowtie L}})$, and

(b) there exist two functions

$$\phi : \mathbb{X}_K \longrightarrow \mathbb{R},$$
$$\psi : \mathbb{X}_L \longrightarrow \mathbb{R},$$

such that

$$\mu^{\downarrow K\cup L}(\boldsymbol{c}) = \phi(\boldsymbol{c}^{\downarrow X_K}) \cdot \psi(\boldsymbol{c}^{\downarrow X_L})$$

for all $\boldsymbol{c} \in 2^{\mathbb{X}_{K\bowtie L}}$.

As we already said above, $\mu_1 \overset{d}{\triangleright} \mu_2$ and $\mu_1 \overset{f}{\triangleright} \mu_2$ generally differ from each other. Nevertheless, there are special situations in which they coincide. For example, it is easy to show that these compositions coincide when $K \cap L = \emptyset$, or when the composed bpas are probabilistic. Nevertheless, note that specification of necessary and sufficient conditions under which the two operators coincide has remained an open problem for several years.

4 Decomposable Models

By decomposable probability distributions we understand the distributions whose conditional dependence structures can be well depicted with the help of so-called *decomposable graphs* [11]. The latter is an important class of graphs that were introduced in graph theory under several different names (triangulated graphs, chordal graphs – see [13]). One of the characteristic properties of these graphs is that their cliques can be ordered to meet the so-called running intersection property (see below).

In general, by decomposable models we understand multidimensional measures/distributions/valuations that can be decomposed, and thereafter reconstructed from a system of its marginals without loss of information, and the structure of the system of marginals can be depicted with the help of a decomposable graph. The latter condition is equivalent to the requirement that the marginals can be ordered to meet the running intersection property. The purpose of such decomposition is twofold. One reason is to decrease the number of necessary parameters representing the multidimensional model. The other reason is to decrease the computational complexity of the procedures that process this model (e.g., when used for inference). As a rule, these two goals are mutually connected. Usually, the fewer parameters necessary to define a model, the more efficient the computational procedures will be.

Let us also apply this general idea to bpas in D-S theory. Consider a multidimensional bpa μ on $\mathbb{X}_N$. Let $\{K_1, K_2, \ldots, K_m\}$ be a cover of N (i.e., $\cup_{i=1}^m K_i = N$) meeting the *running intersection property* (RIP):

$$\forall i = 2, \ldots, m \;\; \exists j < i : \;(K_1 \cup \ldots \cup K_{i-1}) \cap K_i \subseteq K_j.$$

We say that μ is a *decomposable model with structure* $\{K_1, K_2, \ldots, K_m\}$ if

$$\mu = \mu^{\downarrow K_1} \triangleright \mu^{\downarrow K_2} \triangleright \ldots \triangleright \mu^{\downarrow K_m}. \qquad (5)$$

Since the operator of composition is not associative, we must explain how to interpret the right hand side of Formula (5): whenever the order of operators is not specified by parentheses, they are performed from left to right, i.e.,

$$\mu^{\downarrow K_1} \triangleright \mu^{\downarrow K_2} \triangleright \ldots \triangleright \mu^{\downarrow K_m} = \left(\ldots(\mu^{\downarrow K_1} \triangleright \mu^{\downarrow K_2}) \triangleright \ldots \triangleright \mu^{\downarrow K_{m-1}}\right) \triangleright \mu^{\downarrow K_m}.$$

As we used the general symbol $\triangleright$, the reader certainly understands that, in principal, either of the operators ${}^{d}\!\triangleright$ and ${}^{f}\!\triangleright$ can be used. This is why we will use the notions of d-decomposability and f-decomposability when we need to stress which of the two operators of composition is being applied.

The following property of decomposable models in VBS framework was proven in [9], and therefore it also holds for the considered decomposable models in D-S theory: If $K_{j_1}, K_{j_2}, \ldots, K_{j_m}$ is a permutation of $K_1, K_2, \ldots, K_m$ such that it also meets RIP, then

$$\mu^{\downarrow K_1} \triangleright \mu^{\downarrow K_2} \triangleright \ldots \triangleright \mu^{\downarrow K_m} = \mu^{\downarrow K_{j_1}} \triangleright \mu^{\downarrow K_{j_2}} \triangleright \ldots \triangleright \mu^{\downarrow K_{j_m}}.$$

Let us conclude this section by highlighting that in this paper we restrict our attention only to sequential models, i.e., the distributions that can be expressed in the form of Formula (5). For the properties of more general compositional models see [12].

4.1 Marginal Problem Example

Perhaps the best way to illustrate both advantages and problems connected with computations on decomposable models is to consider a real (maximally simplified) task. Let $\mu_1, \ldots, \mu_4$ be four bpas on $\mathbb{X}_{\{1,2\}}$, $\mathbb{X}_{\{2,3\}}$, $\mathbb{X}_{\{3,4\}}$, $\mathbb{X}_{\{1,4\}}$, respectively. The goal is to find a bpa μ^* on $\mathbb{X}_{\{1,2,3,4\}}$ such that all μ_i are its marginals.

To the best of our knowledge, there is no better way to solve this problem than to apply *Iterative Proportional Fitting Procedure* (IPFP) [4, 3], which proceeds as follows:

I Define bpa $\lambda(\mathbf{c}) := \frac{1}{|2^{\mathbb{X}_{\{1,2,3,4\}}}|}$ for all nonempty $\mathbf{c} \subseteq \mathbb{X}_{\{1,2,3,4\}}$

II Repeat the following cycle (four steps) until the procedure converges:

(i) $\lambda := \mu_1 \triangleright \lambda$,

(ii) $\lambda := \mu_2 \triangleright \lambda$,

(iii) $\lambda := \mu_3 \triangleright \lambda$,

(iv) $\lambda := \mu_4 \triangleright \lambda$.

We dealt with this procedure previously in ISIPTA contribution [8] where we showed that

(a) for both operators ${}^{d}\!\triangleright$ and ${}^{f}\!\triangleright$ it holds that if the procedure converges then all bpas $\mu_1, \ldots, \mu_4$ are marginals of the resulting bpa (more precisely of the limit bpa, to which the procedure converges);

(b) if there exists a bpa having all four bpas $\mu_1, \ldots, \mu_4$ for its marginals then the procedure with ${}^{f}\!\triangleright$ converges;

(c) it may happen that the procedure with ${}^{d}\!\triangleright$ does not converge even if there exists a bpa having all four bpas $\mu_1, \ldots, \mu_4$ for its marginals.

These results give a hint that using IPFP with ${}^{f}\!\triangleright$ should be preferred to ${}^{d}\!\triangleright$; in fact, ${}^{f}\!\triangleright$ is computationally less demanding and the convergence of the procedure is guaranteed by the existence of a single join extension of the given marginal.

Nevertheless, let us note that, regardless whether ${}^{d}\!\triangleright$ or ${}^{f}\!\triangleright$ is used, the procedure cannot be applied to multidimensional bpas because at each step we have to compute all $|2^{\mathbb{X}_N}|$ values when computing bpa λ. Even in the considered simple four-dimensional case it means that we have to compute $|2^{\mathbb{X}_{\{1,2,3,4\}}}|$ values (which equals 65 535 for binary, and 43 046 720 for ternary variables), by which bpa λ (or the respective comf, in case $\overset{d}{\triangleright}$ is used) is defined. In analogy to computation within the probabilistic framework [5], a principal simplification can be achieved when representing the computed bpa λ in the form of a decomposable model.

When computing on decomposable models, the general approach starts with finding a RIP cover of N that is a coarsening of $(K_1, K_2, \ldots, K_m)$. In the considered example it means that we look for a RIP cover of $\{1,2,3,4\}$ that is a coarsening of $\{\{1,2\},\{2,3\},\{3,4\},\{1,4\}\}$. This property is met by $\{\{1,2,3\},\{1,3,4\}\}$ (the other possibility would be $\{\{1,2,4\},\{2,3,4\}\}$). Thus, we will consider a decomposable model

$$\lambda = \lambda_1 \triangleright \lambda_2,$$

where $\lambda_1 = \lambda^{\downarrow\{1,2,3\}}$ and $\lambda_2 = \lambda^{\downarrow\{1,3,4\}}$. This type of representation of the four-dimensional pba λ claims a formal change of step II of the above described IPFP algorithm; each step of the cycle is split into two simpler steps – see the modified algorithm below.

However, if we decide to decrease computational complexity of the presented algorithm by the decomposition of the computed bpa, we must be ready to face new problems that do not appear in the probabilistic framework. Namely, in D-S theory the "uniform" bpa (i.e., the initial bpa that is assigned in step I of the algorithm) is *not* decomposable. The reader can see it immediately from the fact that in the considered example, decomposable bpas have only focal elements from $2^{\mathbb{X}_{\{1,2,3\}\bowtie\{2,3,4\}}}$, which contains only 9 999 sets out of 65 535 from $2^{\mathbb{X}_{\{1,2,3,4\}}}$. Therefore, in the considered four-dimensional example, when applying IPFP to the considered decomposable models we also have to modify the initializing step I. So, for the considered example we are getting the following modified algorithm.

I Define bpas:
$$\lambda_1(\mathbf{c}) := \frac{1}{|2^{|\mathbb{X}_{\{1,2,3\}}|}|} \text{ for all } \mathbf{c} \in 2^{\mathbb{X}_{\{1,2,3\}}},$$
$$\lambda_2(\mathbf{c}) := \frac{1}{|2^{|\mathbb{X}_{\{1,3,4\}}|}|} \text{ for all } \mathbf{c} \in 2^{\mathbb{X}_{\{1,3,4\}}}.$$

II Repeat the following cycle (eight steps) until the procedure converges:

(i) $\lambda_1 := \mu_1 \triangleright \lambda_1$,

(ii) $\lambda_2 := \lambda_1^{\{1,3\}} \triangleright \lambda_2$,

(iii) $\lambda_1 := \mu_2 \triangleright \lambda_1$,

(iv) $\lambda_2 := \lambda_1^{\{1,3\}} \triangleright \lambda_2$,

(v) $\lambda_2 := \mu_3 \triangleright \lambda_2$,

(vi) $\lambda_1 := \lambda_2^{\{1,3\}} \triangleright \lambda_1$,

(vii) $\lambda_2 := \mu_4 \triangleright \lambda_2$,

(viii) $\lambda_1 := \lambda_2^{\{1,3\}} \triangleright \lambda_1$.

(Let us note that it does not matter that λ_1, λ_2 defined in the initializing step of the algorithm may be inconsistent. The only condition we have to guarantee is that $\lambda_1 \triangleright \lambda_2$ is positive for all $\mathbf{c} \in 2^{\mathbb{X}_{\{1,2,3\}\bowtie\{2,3,4\}}}$.)

The achieved simplification is obvious. Although we have to perform twice as many steps in one cycle when considering the simple decomposable model as we do in the general case, we only compute 255 numbers at each step for binary variables instead of 65 535. Naturally, the computational savings for ternary variables would be even more progressive. So, at this point, the question remains (it will be discussed in the next section) whether it is preferable to consider d-decomposable or f-decomposable models.

Before leaving the example let us set right one theoretical issue about this modified algorithm. Recalling the results from the last ISIPTA paper, we said (see point (b) above) that the algorithm with $\overset{f}{\triangleright}$ converges if there exists a bpa having all $\mu_1, \ldots, \mu_4$ for its marginals. This holds because the general IPFP algorithm is initialized with bpa λ, which is positive on $2^{\mathbb{X}_{\{1,2,3,4\}}}$, and therefore it dominates all bpas on $\mathbb{X}_{\{1,2,3,4\}}$. It is just a question of going through the proof of the assertion guaranteeing convergence of the IPFP procedure in [8], to show that, for the decomposable version of the IPFP algorithm, only a weaker assertion holds:

(b) if there exists a decomposable bpa (decomposable with structure $\{\{1,2,3\},\{1,3,4\}\}$) having all four bpas $\mu_1, \ldots, \mu_4$ for its marginals, then the procedure with $\overset{f}{\triangleright}$ converges;

5 Inference

The simplest inference scenario is based on computation of conditionals. Instructions for computing conditionals (and a clarification of what their properties are) can be found in [9]. In that paper it was shown that, for bpa μ on $\mathbb{X}_M$ and $X_j, X_k \in M$,

$$\mu(X_k | X_j = a) = \left(\nu_{X_j=a} \overset{d}{\triangleright} \mu\right)^{\downarrow X_k},$$

where $\nu_{X_j=a}$ is a one-dimensional bpa on $\mathbb{X}_j$ having just one focal element $\{a\} \subset \mathbb{X}_j$, for which $\nu_{X_j=a}(\{a\}) = 1$. (Note that $\nu_{X_j=a}$ is thus normal and proper.) This bpa expresses the fact that we are sure that variable X_j achieves value a.

Maybe it is worth showing that for computation of conditional bpas we *must* use the operator ${}^{d}\!\triangleright$ and not ${}^{f}\!\triangleright$.

5.1 Example

Consider two variables X_1, X_2 with $\mathbb{X}_1 = \mathbb{X}_2 = \{b, c, d, e\}$, and bpa μ with just one focal element

$$\mu(\{(b,b),(c,c),(d,d),(e,e)\}) = 1.$$

This bpa describes the situation when we do not know which of the values occurs, but we are sure that both variables X_1 and X_2 certainly have the same value.

To compute $\nu_{X_1=b} \,{}^{d}\!\triangleright\, \mu$, we proceed according to Definition 2. First notice that $\nu_{X_1=b}$ is a probabilistic bpa, and therefore (see Formula (3)) it is the same as the corresponding comf $\theta_{X_1=b} = \nu_{X_1=b}$. Denote by θ (with no index) the comf corresponding to μ. The marginal $\mu^{\downarrow X_1}$ is a vacuous bpa with just one focal element $\mathbb{X}_1 = \{b, c, d, e\}$. Therefore, the corresponding comf $\theta^{\downarrow X_1}$ equals 1 for all nonempty subsets of $\mathbb{X}_1$, and therefore the denominator in the formula appearing in Definition 2 equals 1. So, in this specific case we get

$$\theta_{X_1=b} \triangleright \theta = \theta_{X_1=b} \cdot \theta.$$

Due to Formula (3), θ equals 1 for all nonempty subsets of $\{(b,b),(c,c),(d,d),(e,e)\}$. Therefore $\theta_{X_1=b} \,{}^{d}\!\triangleright\, \theta$ equals one for all those subsets $\mathbf{a} \subseteq \{(b,b),(c,c),(d,d),(e,e)\}$ in which no other value than b appears at the first position, i.e., $\mathbf{a}^{\downarrow\{X_1\}} = \{b\}$, and it is only $\{(b,b)\}$. Therefore we get

$$(\theta_{X_1=b} \,{}^{d}\!\triangleright\, \theta)(\{(b,b)\}) = 1,$$

and $(\theta_{X_1=b} \,{}^{d}\!\triangleright\, \theta)(\mathbf{a}) = 0$ for all $\mathbf{a} \subseteq \mathbb{X}_1 \times \mathbb{X}_2$, for which $\mathbf{a} \neq \{(b,b)\}$.

This means that we get a probabilistic bpa equaling 1 for $\{(b,b)\}$, from which we get (after marginalization) that $\mu(X_2|X_1 = b)$ equals 1 if and only if $X_2 = b$.

However, if we computed $\mu(X_2|X_1 = b) = \nu_{X_1=b} \,{}^{f}\!\triangleright\, \mu$ according to Definition 3 we would get that $\nu_{X_1=b} \,{}^{f}\!\triangleright\, \mu$ has, again, only one focal element, but this time it would be $\{(b,b),(b,c),(b,d),(b,e)\}$. Therefore, marginalizing this bpa for variable X_2 we would get a vacuous bpa for which

$$(\nu_{X_1=b} \,{}^{f}\!\triangleright\, \mu)^{\downarrow X_2}(\{b,c,d,e\}) = 1.$$

This equality does not correspond to what we expected.

5.2 Conditioning in Decomposable Models

Consider decomposable bpa μ with the structure $\{K_1, K_2, \ldots, K_m\}$ ($\bigcup_{i=1}^m K_i = N$), and assume, first, that it is a d-decomposable model, i.e.,

$$\mu = \mu^{\downarrow K_1} \,{}^{d}\!\triangleright\, \mu^{\downarrow K_2} \,{}^{d}\!\triangleright\, \ldots \,{}^{d}\!\triangleright\, \mu^{\downarrow K_m}.$$

If we want to compute a conditional

$$\begin{aligned} &\mu(X_k|X_j = a) \\ &\quad = \left(\nu_{X_j=a} \,{}^{d}\!\triangleright\, (\mu^{\downarrow K_1} \,{}^{d}\!\triangleright\, \mu^{\downarrow K_2} \,{}^{d}\!\triangleright\, \ldots \,{}^{d}\!\triangleright\, \mu^{\downarrow K_m})\right)^{\downarrow X_k} \end{aligned}$$

we can be facing a computationally hard problem, unless we take into account the fact that $K_1, K_2, \ldots, K_m$ are ordered to meet RIP. This enables us to carry out the necessary computations *locally* in the way that was shown in [7]. This computationally tractable process takes advantage of the well-known fact (an immediate consequence of the existence of a join tree, see [1]) that if $K_1, K_2, \ldots, K_m$ can be ordered to meet RIP, then for each $\ell \in \{1, 2, \ldots, m\}$ there exists an ordering that meets RIP and in which K_ℓ is the first one. So consider any K_ℓ for which $X_j \in K_\ell$, and find the ordering that meets RIP and starts with K_ℓ. Without loss of generality, let it be $K_1, K_2, \ldots, K_m$ (so, in this case we assume that $X_j \in K_1$). This fact makes the application of Property 8 of Theorem 1 possible; applying it $(m-1)$ times we get

$$\begin{aligned} &\nu_{X_j=a} \,{}^{d}\!\triangleright\, (\mu^{\downarrow K_1} \,{}^{d}\!\triangleright\, \mu^{\downarrow K_2} \,{}^{d}\!\triangleright\, \ldots \,{}^{d}\!\triangleright\, \mu^{\downarrow K_m}) \\ &\quad = \nu_{X_j=a} \,{}^{d}\!\triangleright\, (\mu^{\downarrow K_1} \,{}^{d}\!\triangleright\, \mu^{\downarrow K_2} \,{}^{d}\!\triangleright\, \ldots \,{}^{d}\!\triangleright\, \mu^{\downarrow K_{m-1}}) \\ &\qquad\qquad {}^{d}\!\triangleright\, \mu^{\downarrow K_m} = \ldots \\ &\quad = \nu_{X_j=a} \,{}^{d}\!\triangleright\, \mu^{\downarrow K_1} \,{}^{d}\!\triangleright\, \mu^{\downarrow K_2} \,{}^{d}\!\triangleright\, \ldots \,{}^{d}\!\triangleright\, \mu^{\downarrow K_m} \end{aligned}$$

Notice that the whole process can be repeated several times in case one wants to compute a conditional like $\mu(X_k|X_1 = a, X_3 = b, \ldots)$ – the only additional problem is that before each reordering of marginals (i.e., after each conditioning by one variable) one has to ensure that the model is composed from marginals of one bpa (This can be done by a simple computational process described in Proposition 7 of [7]). So, we can conclude that the computation of conditionals in d-decomposable models can be carried out locally.

Another question is whether the same computationally local process can be also used if we consider f-decomposable model

$$\mu = \mu^{\downarrow K_1} \,{}^{f}\!\triangleright\, \mu^{\downarrow K_2} \,{}^{f}\!\triangleright\, \ldots \,{}^{f}\!\triangleright\, \mu^{\downarrow K_m}.$$

However, though we conjecture that the answer is positive, at this moment the question still remains open. We would need to prove the following assertion to confirm the validity of our expectation.

Conjecture Suppose μ_1, μ_2 and μ_3 are bpas on $\mathbb{X}_K$, $\mathbb{X}_L$, and $\mathbb{X}_M$, respectively. If $L \supset (K \cap M)$ then, $(\mu_1 \,{}^{d}\!\triangleright\, \mu_2) \,{}^{f}\!\triangleright\, \mu_3 = \mu_1 \,{}^{d}\!\triangleright\, (\mu_2 \,{}^{f}\!\triangleright\, \mu_3)$.

6 Summary and Conclusions

The primary goal of this paper is to convince the reader that introducing two operators of composition for belief functions is not an end in itself. Each of them has its own raison d'être. $\overset{d}{\triangleright}$ is, in a way, a generalization of probabilistic composition, introducing a conditional independence among the variables, whereas $\overset{f}{\triangleright}$ generalizes probabilistic factorization. Since these two notions coincide in probability theory, it is sufficient to use just one operator of composition in probability theory.

The role of $\overset{d}{\triangleright}$ for computation of conditionals is irreplaceable. On the other hand, computational procedures for f-decomposable models are much more efficient than those for d-decomposable models. The only problem spoiling their mutually advantageous coexistence will disappear once the presented conjecture is proven. Nevertheless, even if the conjecture is disproved there will still be a chance to design computationally efficient procedures employing both studied operators; they just will not be as simple as the procedure described in the last section.

Acknowledgements

This work was partially supported by GAČR under Grant No. 15-00215S.

References

[1] C. Beeri, R. Fagin, D. Maier, and M. Yannakakis. On the Desirability of Acyclic Database Schemes. *J. ACM*, 30, 3, pp. 479–513, 1983.

[2] B. Ben Yaghlane, Ph. Smets, and K. Mellouli. Belief functions independence: II. the conditional case. *Int. J. Approx. Reasoning*, 31, pp. 31–75, 2002.

[3] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.* 3, pp 146–158, 1975.

[4] W. E. Deming and F. F. Stephan. On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Ann. Math. Statist.* 11, 427–444, 1940.

[5] R. Jiroušek. Solution of the Marginal Problem and Decomposable Distributions. *Kybernetika* 27, 5, pp. 403–412, 1991.

[6] R. Jiroušek. Foundations of compositional model theory. *Int. J. General Systems*, 40, 6, pp 623–678, 2011.

[7] R. Jiroušek. Local computations in Dempster-Shafer theory of evidence. *Int. J. Approx. Reasoning*, 53, 8, pp. 1155–1167, 2012.

[8] R. Jiroušek and V. Kratochvíl. On Open Problems Connected with Application of the Iterative Proportional Fitting Procedure to Belief Functions. In *Proc. of the 8th Symp. on Imprecise Probabilities and Their Applications,* F. Cozman, T. Denœux, S. Desterecke, T. Seidenfeld, Eds., Compiègne, France, pp. 149–158, 2013.

[9] R. Jiroušek and P. P. Shenoy. Compositional models in valuation-based systems. *Int. J. Approx. Reasoning*, 55, 1, pp. 277–293, 2014.

[10] R. Jiroušek, J. Vejnarová and M. Daniel. Compositional models of belief functions. In *Proc. of the 5th Symp. on Imprecise Probabilities and Their Applications,* G. de Cooman, J. Vejnarová, M. Zaffalon, Eds., Praha, pp. 243–252, 2007.

[11] S. L. Lauritzen, *Graphical models.* Oxford University Press, 1996.

[12] F. Malvestuto. Equivalence of compositional expressions and independence relations in compositional models. *Kybernetika*, 50, 3, pp. 322-362, 2014.

[13] R. Merris. *Graph theory.* John Wiley, New York, 2001.

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

[15] P. P. Shenoy. Conditional independence in valuation-based systems. *Int. J. Approx. Reasoning,* 10, 3, pp. 203–234, 1994.

[16] J. Vejnarová. On conditional independence in evidence theory. In *Proc. of the 6th Symposium on Imprecise Probability: Theories and Applications,* Durham, UK, pp. 431–440, 2009.