# Factorization and Decomposable Models in Dempster-Shafer Theory of Evidence

Radim Jiroušek
Faculty of Management, University of Economics
Institute of Information Theory and Automation, Academy of Sciences
Jindřichův Hradec, Czech Republic
Email: radim@utia.cas.cz

***Abstract*—The paper presents our attempt to define the concept of factorization within the Dempster-Shafer theory of evidence. In the same way as in probability theory, the introduced concept can support procedures for efficient multidimensional model construction and processing. The main result of this paper is a *factorization lemma* describing, in the same way as in probability theory, the relationship between factorization and conditional independence.**

**Keywords: Discrete belief functions, conditional independence, multidimensional model.**

## I. Introduction

It is well-known that problems of practice require knowledge bases comprising great number of attributes/properties. Therefore, when considering probabilistic or possibilistic models one has to cope with distributions of hundreds or even thousands of dimensions. The dimensionality explosion starts to be almost hopeless when considering models of Dempster-Shafer theory [4], [9], where the basic assignment is not a point function, like distributions in probability/possibility theories, but a set function. This is why any space-saving technique for model representation and/or processing [11] is in Dempster-Shafer theory of such a great importance.

Speaking about probability theory, it became widely known that a substantial decrease of computational complexity was achieved with the help of models taking advantage of the concept of conditional independence. However, studying properly probabilistic graphical Markov models one can realize that it is not the notion of *conditional independence* that makes it possible to represent these models efficiently. The efficiency is based on a *factorization*, which in probability theory (due to *factorization lemma* presented here as Lemma 1) coincide with the conditional independence. Going into details, one can notice that the notion of factorization has been introduced in several different ways in probability theory. The goal of this paper is to show that factorization can be exploited also in Dempster-Shafer theory. Before doing this, we briefly analyze the notion of factorization in probability theory.

## II. Probabilistic factorization

In this section we will introduce a necessary notation and recall several notions from probability theory, which served as an inspiration for the considerations presented in the further parts of this paper. Here, we will consider a probability measure $\pi$ on a finite space

$$\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_n,$$

i.e. an additive set function

$$\pi : \mathcal{P}(\mathbf{X}_N) \longrightarrow [0,1],$$

for which $\pi(\mathbf{X}_N) = 1$. For any $K \subseteq N$, symbol $\pi^{\downarrow K}$ will denote its respective marginal measure (for each $B \subseteq \mathbf{X}_K$):

$$\pi^{\downarrow K}(B) = \sum_{\substack{A \subseteq \mathbf{X}_N \\ A^{\downarrow K} = B}} \pi(A),$$

which is a probability measure on subspace

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

Let us remark that for $K = \emptyset$ we get $\pi^{\downarrow \emptyset} = 1$. Analogous notation will be used also for projections of points and sets. For a point $x = (x_1, x_2, \ldots, x_n) \in \mathbf{X}_N$ its projection into subspace $\mathbf{X}_K$ will be denoted

$$x^{\downarrow K} = (x_{i, i \in K}),$$

and for $A \subseteq \mathbf{X}_N$

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

Consider a probability measure $\pi$ and three disjoint groups of variables $X_K = \{X_i\}_{i \in K}$, $X_L = \{X_i\}_{i \in L}$ and $X_M = \{X_i\}_{i \in M}$ ($K, L, M \subset N$, $K \neq \emptyset \neq L$). We say that $X_K$ and $X_L$ are *conditionally independent given* $X_M$ (in probability measure $\pi$) if for all[2] $x \in \mathbf{X}_{K \cup L \cup M}$

$$\pi^{\downarrow K \cup L \cup M}(x) \cdot \pi^{\downarrow M}(x^{\downarrow M}) = \pi^{\downarrow K \cup M}(x^{\downarrow K \cup M}) \cdot \pi^{\downarrow L \cup M}(x^{\downarrow L \cup M}).$$

This property will be denoted by the symbol $K \perp\!\!\!\perp L \,|\, M \; [\pi]$. In case that $M = \emptyset$ then we say that groups of variables $X_K$ and $X_L$ are (*unconditionally*[3]) *independent*, which is usually denoted by a simplified notation: $K \perp\!\!\!\perp L \; [\pi]$.

[1]This work was supported by GAČR under the grant no. ICC/08/E010, and 201/09/1891, and by MŠMT ČR under grants 1M0572 and 2C06019.

[2]As usually, we do not distinguish between a singleton set and its element, so $x$ stands also for $\{x\}$, and $x^{\downarrow K}$ is the only element from $\{x\}^{\downarrow K}$.

[3]Some authors call it marginal independence.

As already mentioned above, the notion of factorization is introduced in probability theory in several different ways, and therefore we will use some adjectives to distinguish them from each other. The properties presented further in this section as lemmata and corollaries can either be found in [8] or can be directly deduced as trivial consequences of the properties presented there.

*Simple factorization:* Consider two nonempty sets $K, L \subseteq N$. We say that $\pi$ *factorizes with respect to* $(K, L)$ if there exist two nonnegative functions

$$\phi : \mathbf{X}_K \longrightarrow [0, +\infty) \quad \text{and} \quad \psi : \mathbf{X}_L \longrightarrow [0, +\infty),$$

such that for each $x \in \mathbf{X}_{K\cup L}$ the equality

$$\pi^{\downarrow K\cup L}(x) = \phi(x^{\downarrow K}) \cdot \psi(x^{\downarrow L})$$

holds true.

*Lemma 1 (Factorization lemma):* Let $K, L \subseteq N$ be nonempty. $\pi$ factorizes with respect to $(K, L)$ if and only if $K \setminus L \perp\!\!\!\perp L \setminus K | K \cap L\ [\pi]$.

*Corollary 1:* $\pi$ factorizes with respect to $(K, L)$ if and only if

$$\pi^{\downarrow K\cup L}(x) = \pi^{\downarrow K}(x^{\downarrow K}) \cdot \pi^{\downarrow K\cup L}(x^{\downarrow L\setminus K}|x^{\downarrow K\cap L}),$$

for all $x \in \mathbf{X}_{K\cup L}$.

*Corollary 2:* Let $\pi_1, \pi_2, \pi_3, \ldots$ be a sequence of probability measures each of them factorizing with respect to $(K, L)$. If this sequence is convergent then also the limit measure $\lim_{j\to+\infty} \pi_j$ factorizes with respect to $(K, L)$.

*Multiple factorization:* Consider a finite system of nonempty subsets $K_1, K_2, \ldots, K_r$ of a set $N$. We say that $\pi$ *factorizes with respect to* $(K_1, K_2, \ldots, K_r)$ if there exist $r$ functions ($i = 1, 2, \ldots, r$)

$$\phi_i : \mathbf{X}_{K_i} \longrightarrow [0, +\infty),$$

such that for all[4] $x \in \mathbf{X}_{K_1\cup\ldots\cup K_r}$

$$\pi^{\downarrow K_1\cup\ldots\cup K_r}(x) = \prod_{i=1}^{r} \phi_i(x^{\downarrow K_i}).$$

*Remark:* In this general case one can (using Lemma 1) derive a system of conditional independence relations valid for a measure $\pi$ factorizing with respect to $(K_1, K_2, \ldots, K_r)$ but no assertion that could be considered a direct analogy to any of the preceding Corollaries holds true. This is why the following type of factorization is often considered.

[4] Sometimes, the validity of this equality is required only for those points $x$ for which $\pi^{\downarrow K_1\cup\ldots\cup K_r}(x) > 0$.

*Recursive factorization:* Consider a finite system of nonempty subsets $K_1, K_2, \ldots, K_r$ of a set $N$. We say that $\pi$ *recursively factorizes with respect to* $(K_1, K_2, \ldots, K_r)$ if for each $i = 2, \ldots, r$ $\pi$ (simply) factorizes with respect to the pair $(K_1 \cup \ldots \cup K_{i-1}, K_i)$.

*Remark:* Using Corollary 1 iteratively one can get a formula expressing the multidimensional measure $\pi^{\downarrow K_1\cup\ldots\cup K_r}$ with the help of its respective marginals[5]

$$\pi^{\downarrow K_1\cup\ldots\cup K_r}(x) = \prod_{i=1}^{r} \pi^{\downarrow K_i}(x^{\downarrow K_i\setminus(K_1\cup\ldots\cup K_{i-1})}|x^{\downarrow K_i\cap(K_1\cup\ldots\cup K_{i-1})}).$$

So we are getting a trivial assertion saying that if $\pi$ recursively factorizes with respect to $K_1, K_2, \ldots, K_r$ then it also factorizes with respect to this system of subsets. Let us stress that recursive factorization is much stronger than multiple factorization. For example, for recursive factorization an analogy to Corollary 2 holds true.

*Decomposition:* We say that a sequence $K_1, K_2, \ldots, K_r$ meets the *running intersection property* (RIP) if for all $i = 2, \ldots, r$ there exists $j, 1 \leq j < i$, such that

$$K_i \cap (K_1 \cup \ldots \cup K_{i-1}) \subseteq K_j.$$

In case of factorization with respect to a sequence of sets meeting RIP we are getting the strongest type of factorization. Namely, in this case multiple and recursive factorizations coincide.

*Lemma 2 (Decomposition lemma):* If $(K_1, K_2, \ldots, K_r)$ meets RIP, then $\pi$ factorizes with respect to $(K_1, K_2, \ldots, K_r)$ if and only if it recursively factorizes with respect to $(K_1, K_2, \ldots, K_r)$.

In the literature (e.g. in [8]), measures factorizing with respect to systems of sets meeting (after a possible reordering) RIP are usually called decomposable measures. In this text we will say that they are *decomposable with respect to* $(K_1, K_2, \ldots, K_r)$. As a direct consequence of what have been said before, one can see that a distribution decomposable with respect to sequence $(K_1, K_2, \ldots, K_r)$ can be expressed as a product of its conditional marginals

$$\pi^{\downarrow K_1\cup\ldots\cup K_r}(x) = \prod_{i=1}^{r} \pi^{\downarrow K_i}(x^{\downarrow K_i\setminus(K_1\cup\ldots\cup K_{i-1})}|x^{\downarrow K_i\cap(K_1\cup\ldots\cup K_{i-1})}),$$

and that the limit a convergent sequence of measures decomposable with respect to $(K_1, K_2, \ldots, K_r)$ is also decomposable with respect to the same sequence.

*Remark:* It is important to realize that since two sets $(K, L)$ always meet RIP, all the presented definitions coincide if one considers factorization with respect to only two sets. This is

[5] Read $(K_1 \cup \ldots \cup K_0)$ as $\emptyset$.

perhaps why not many authors distinguish different types of factorization.

*Remark:* The presented list of types of factorization is not comprehensive. For example, one can consider a marginal factorization requiring that the respective measure is uniquely given by a system of its marginals; for example as a maximum entropy extension.

## III. Dempster-Shafer theory - notation

As in the previous section, we consider a finite *multidimensional frame of discernment*

$$\mathbf{X}_N = \mathbf{X}_1 \times \mathbf{X}_2 \times \ldots \times \mathbf{X}_n,$$

and its *subframes* $\mathbf{X}_K$. Consider $K, L \subseteq N$ and $M \subseteq K$. In addition to a projection of a set $A$ we will need also an opposite operation, which will be called a join. By a *join* of two sets $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_L$ we will understand a set

$$A \otimes B = \{x \in \mathbf{X}_{K\cup L} : x^{\downarrow K} \in A \;\;\&\;\; x^{\downarrow L} \in B\}.$$

Let us note that if $K$ and $L$ are disjoint, then $A \otimes B = A \times B$, if $K = L$ then $A \otimes B = A \cap B$.

In view of this paper it is important to realize that if $x \in C \subseteq \mathbf{X}_{K\cup L}$, then $x^{\downarrow K} \in C^{\downarrow K}$ and $x^{\downarrow L} \in C^{\downarrow L}$, which means that always

$$C \subseteq C^{\downarrow K} \otimes C^{\downarrow L}.$$

However, and it is of great importance in this paper, it does not mean that $C = C^{\downarrow K} \otimes C^{\downarrow L}$. For example, considering 3-dimensional frame of discernment $\mathbf{X}_{\{1,2,3\}}$ with $\mathbf{X}_i = \{a_i, \bar{a}_i\}$ for all three $i = 1, 2, 3$, and $C = \{a_1a_2a_3, \bar{a}_1a_2a_3, a_1a_2\bar{a}_3\}$ one gets

$$\begin{aligned} C^{\downarrow\{1,2\}} \otimes C^{\downarrow\{2,3\}} &= \{a_1a_2, \bar{a}_1a_2\} \otimes \{a_2a_3, a_2\bar{a}_3\} \\ &= \{a_1a_2a_3, \bar{a}_1a_2a_3, a_1a_2\bar{a}_3, \bar{a}_1a_2\bar{a}_3\} \\ &\supsetneq C. \end{aligned}$$

In Dempster-Shafer theory of evidence several measures are used to model the uncertainty (belief, plausibility and commonality measures). All of them can be defined with the help of another set function called a *basic (probability* or *belief) assignment* $m$ on $\mathbf{X}_K$, i.e.

$$m : \mathcal{P}(\mathbf{X}_K) \longrightarrow [0, 1]$$

for which $\sum_{A\subseteq \mathbf{X}_K} m(A) = 1$. Since we will consider in this paper only normalized basic assignments we will assume that $m(\emptyset) = 0$. Set $A \subseteq \mathbf{X}_K$ is said to be a *focal element* of $m$ if $m(A) > 0$.

Analogously to marginal probability measures, we consider also marginal basic assignments of $m$ defined on $\mathbf{X}_N$. For each $K \subseteq N$ a *marginal basic assignment* of $m$ is defined (for each $B \subseteq \mathbf{X}_K$):

$$m^{\downarrow K}(B) = \sum_{\substack{A\subseteq \mathbf{X}_N \\ A^{\downarrow K}=B}} m(A).$$

Considering two basic assignments $m_1$ and $m_2$ defined on $\mathbf{X}_K$ and $\mathbf{X}_L$, respectively, we say that they are *projective* if

$$m_1^{\downarrow K\cap L} = m_2^{\downarrow K\cap L}.$$

## IV. Independence and factorization in Dempster-Shafer theory

Let us now present a generally accepted notion of independence ( [1], [10], [12])[6].

*Definition 1:* Let $m$ be a basic assignment on $\mathbf{X}_N$ and $K, L \subset N$ be nonempty disjoint. We say that groups of variables $X_K$ and $X_L$ are *independent*[7] with respect to basic assignment $m$ (in notation $K \perp\!\!\!\perp L\ [m]$) if for all $A \subseteq \mathbf{X}_{K\cup L}$

$$m^{\downarrow K\cup L}(A) = \begin{cases} m^{\downarrow K}(A^{\downarrow K}) \cdot m^{\downarrow L}(A^{\downarrow L}) & \text{if } A = A^{\downarrow K} \times A^{\downarrow L}, \\ 0 & \text{otherwise.} \end{cases}$$

There are several generalizations of this notion of independence corresponding to conditional independence (see for example papers [2], [3], [7], [10], [12]). In this text we will use the generalization, which was introduced in [5][8] and [6], and which differs from the notion of conditional independence used, for example, by Shenoy [10] and Studený [12] (and which is the same as the *conditional non-interactivity* used by Ben Yaghlane *et al.* in [2]).

*Definition 2:* Let $m$ be a basic assignment on $\mathbf{X}_N$ and $K, L, M \subset N$ be disjoint, $K \neq \emptyset \neq L$. We say that groups of variables $X_K$ and $X_L$ are *conditionally independent given* $X_M$ *with respect to* $m$ (and denote it by $K \perp\!\!\!\perp L|M\ [m]$), if for any $A \subseteq \mathbf{X}_{K\cup L\cup M}$ such that $A = A^{\downarrow K\cup M} \otimes A^{\downarrow L\cup M}$ the equality

$$\begin{aligned} & m^{\downarrow K\cup L\cup M}(A) \cdot m^{\downarrow M}(A^{\downarrow M}) \\ &\qquad = m^{\downarrow K\cup M}(A^{\downarrow K\cup M}) \cdot m^{\downarrow L\cup M}(A^{\downarrow L\cup M}) \end{aligned}$$

holds true, and $m^{\downarrow K\cup L\cup M}(A) = 0$ for all the remaining $A \subseteq \mathbf{X}_{K\cup L\cup M}$, for which $A \neq A^{\downarrow K\cup M} \otimes A^{\downarrow L\cup M}$.

Notice that for $M = \emptyset$ the concept coincides with Definition 1 (and therefore also with the definition used in [2], [10], [12]). In addition to this, it was proven in [6] that this notion meets all the properties required from the notion of conditional independence, so-called *semigraphoid properties* ( [8], [12], [13]):

(A1) $K \perp\!\!\!\perp L \,|\, M\,[m] \implies L \perp\!\!\!\perp K \,|\, M\,[m]$

(A2) $K \perp\!\!\!\perp L \cup M \,|\, J\,[m] \implies K \perp\!\!\!\perp M \,|\, J\,[m]$

(A3) $K \perp\!\!\!\perp L \cup M \,|\, J\,[m] \implies K \perp\!\!\!\perp L \,|\, M \cup J\,[m]$

(A4) $(K \perp\!\!\!\perp L \,|\, M \cup J\,[m])\ \&\ (K \perp\!\!\!\perp M \,|\, J\,[m]) \implies K \perp\!\!\!\perp L \cup M \,|\, J\,[m]$

It should be highlighted here that all these properties (both the semigraphoid properties and the fact that the notion is a generalization of the unconditional independence) hold

[6] Notice, however, that the presented definition is not standard. It was proved in [6] that this definition is equivalent to that used by most of authors, who use the definition based on application of conjunctive combination rule (non-normalized Dempster's rule of combination) [1], [10], or commonality function [12].

[7] Couso et al. [3] call this independence *independence in random sets*, Klir [7] (*non-interactivity)).*

[8] In this paper the notion was called conditional irrelevance.

true also for the notion employed in [2], [10], [12]). The advantage of our term defined in Definition 2 is that it enables us to prove (among others also) Factorization lemma in Dempster-Shafer theory of evidence, which is presented below as Theorem.

Before introducing the definition of simple factorization, let us illustrate the main difference between factorization in probability and Dempster-Shafer theories with the help of a simple example.

*Example:* Consider just 2-dimensional frame of discernment

$$\mathbf{X}_1 \times \mathbf{X}_2 = \{a_1, \bar{a}_1, \mathring{a}_1\} \times \{a_2, \bar{a}_2, \mathring{a}_2\}.$$

Probability measure $\pi$ on $\mathbf{X}_1 \times \mathbf{X}_2$ factorizes if there exist functions $\phi$ and $\psi$ such that

$$\pi(x_1, x_2) = \phi(x_1) \cdot \psi(x_2).$$

It means that to define 9 probabilities[9] of the 2-dimensional measure $\pi$ one has to give 3 values of function $\phi$ and 3 values of function $\psi$ ($3 \times 3 = 9$). However, considering Dempster-Shafer theory, there are $2^9 - 1 = 511$ non-empty subsets of $\mathbf{X}_1 \times \mathbf{X}_2$, and therefore a basic assignment should be defined by[10] 511 numbers. In this case, however, both factor functions $\phi$ and $\psi$, being set functions on $\mathbf{X}_1$ and $\mathbf{X}_2$, respectively, are defined with 7 numbers ($\mathbf{X}_i$ has 7 nonempty subsets). From this, one can immediately see that some of the values of the resulting join basic assignment must be defined in another way than just product of factors. Nevertheless, this requirement is in harmony with the property we expect from a factorizing basic assignment. In probability theory, any probability measure factorizing on $\mathbf{X}_1 \times \mathbf{X}_2$ is a product of two independent 1-dimensional measures. In Dempster-Shafer theory, any basic assignment $m$, for which $1 \perp\!\!\!\perp 2\ [m]$, has at most 49 focal elements, because there are 462 subsets $A$ of $\mathbf{X}_1 \times \mathbf{X}_2$, for which $A \neq A^{\downarrow\{1\}} \times A^{\downarrow\{2\}}$ (and therefore $m(A) = 0$ for these $A$, due to Definition 1).

Generalizing the ideas from this example to overlapping sets of indices we came up with the following notion of simple factorization in Dempster-Shafer theory of evidence.

*Definition 3 (Simple factorization):* Consider two nonempty sets $K, L \subseteq N$. We say that basic assignment $m$ *factorizes with respect to* $(K, L)$ if there exist two nonnegative set functions

$$\phi : \mathcal{P}(\mathbf{X}_K) \longrightarrow [0, +\infty), \quad \psi : \mathcal{P}(\mathbf{X}_L) \longrightarrow [0, +\infty),$$

such that for all $A \subseteq \mathbf{X}_{K \cup L}$

$$m^{\downarrow K \cup L}(A) = \begin{cases} \phi(A^{\downarrow K}) \cdot \psi(A^{\downarrow L}) & \text{if } A = A^{\downarrow K} \otimes A^{\downarrow L} \\ 0 & \text{otherwise.} \end{cases}$$

[9] More exactly: it is enough to determine 8 probabilities; the last can be computed because all of them must sum up to 1.

[10] Again, since we consider only normalized basic assignments one value can be computed from the remaining 510 numbers.

*Theorem (Factorization lemma):* Let $K, L \subseteq N$ be nonempty. $m$ factorizes with respect to $(K, L)$ if and only if

$$K \setminus L \perp\!\!\!\perp L \setminus K \mid K \cap L\ [m].$$

*Proof:* First notice that for $A \subset \mathbf{X}_{K \cup L}$, for which $A \neq A^{\downarrow K} \otimes A^{\downarrow L}$, $m(A) = 0$ in both situations: when $m$ factorizes with respect to $(K, L)$ and when $K \setminus L \perp\!\!\!\perp L \setminus K \mid K \cap L\ [m]$. So, proving implication

$$\begin{aligned} &K \setminus L \perp\!\!\!\perp L \setminus K \mid K \cap L\ [m] \\ &\qquad \Longrightarrow \quad m \text{ factorizes with respect to } (K, L) \end{aligned}$$

is trivial. It is enough to take

$$\phi(A) = m^{\downarrow L}(A),$$

$$\psi(B) = \begin{cases} \frac{m^{\downarrow L}(B)}{m^{\downarrow K \cap L}(B^{\downarrow K \cap L})} & \text{if } m^{\downarrow K \cap L}(B^{\downarrow K \cap L}) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

for all $A \subseteq \mathbf{X}_K$ and $B \subseteq \mathbf{X}_L$.

To prove the opposite implication consider two functions $\phi$ and $\psi$ meeting the properties required by Definition 3, and consider an arbitrary $A \subset \mathbf{X}_{K \cup L}$, for which $A = A^{\downarrow K} \otimes A^{\downarrow L}$. Before we start computing the necessary marginal basic assignments let us realize that

$$\begin{aligned} &\{B \subseteq \mathbf{X}_{K \cup L} : (B = B^{\downarrow K} \otimes B^{\downarrow L}) \& (B^{\downarrow K} = A^{\downarrow K})\} \\ &\qquad = \{A^{\downarrow K} \otimes C : (C \subseteq \mathbf{X}_L) \& (C^{\downarrow K \cap L} = A^{\downarrow K \cap L})\}. \end{aligned}$$

When computing

$$m^{\downarrow K}(A^{\downarrow K}) = \sum_{\substack{B \subseteq \mathbf{X}_{K \cup L} \\ B^{\downarrow K} = A^{\downarrow K}}} m^{\downarrow K \cup L}(B)$$

we can sum up over only those $B$, for which $B = B^{\downarrow K} \otimes B^{\downarrow L}$, because if $B \neq B^{\downarrow K} \otimes B^{\downarrow L}$, as it follows from Definition 3, $m(B) = 0$. So we get

$$\begin{aligned} m^{\downarrow K}(A^{\downarrow K}) &= \sum_{\substack{B \subseteq \mathbf{X}_{K \cup L} \\ B^{\downarrow K} = A^{\downarrow K}}} m^{\downarrow K \cup L}(B) \\ &= \sum_{\substack{B \subseteq \mathbf{X}_{K \cup L} \\ B^{\downarrow K} = A^{\downarrow K} \\ B = B^{\downarrow K} \otimes B^{\downarrow L}}} \phi(B^{\downarrow K}) \cdot \psi(B^{\downarrow L}) \\ &= \sum_{\substack{C \subseteq \mathbf{X}_L \\ C^{\downarrow K \cap L} = A^{\downarrow K \cap L}}} \phi(A^{\downarrow K}) \cdot \psi(C) \\ &= \phi(A^{\downarrow K}) \cdot \sum_{\substack{C \subseteq \mathbf{X}_L \\ C^{\downarrow K \cap L} = A^{\downarrow K \cap L}}} \psi(C). \end{aligned}$$

Computing analogously $m^{\downarrow L}(A^{\downarrow L})$ one gets

$$m^{\downarrow L}(A^{\downarrow L}) = \psi(A^{\downarrow L}) \cdot \sum_{\substack{D \subseteq \mathbf{X}_K \\ D^{\downarrow K \cap L} = A^{\downarrow K \cap L}}} \phi(D).$$

Now, we have to compute $m^{\downarrow K\cap L}(A^{\downarrow K\cap L})$. For this, realize again that

$$\begin{aligned}&\{B \subseteq \mathbf{X}_{K\cup L} : (B = B^{\downarrow K} \otimes B^{\downarrow L}) \& (B^{\downarrow K\cap L} = A^{\downarrow K\cap L})\}\\ &\quad = \{D \otimes C : (D \subseteq \mathbf{X}_K) \& (C \subseteq \mathbf{X}_L)\\ &\qquad\qquad \& (D^{\downarrow K\cap L} = C^{\downarrow K\cap L} = A^{\downarrow K\cap L})\}.\end{aligned}$$

Using this we get

$$\begin{aligned}&m^{\downarrow K\cap L}(A^{\downarrow K\cap L})\\ &\quad = \sum_{\substack{B\subseteq \mathbf{X}_{K\cup L}\\ B^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} m^{\downarrow K\cup L}(B)\\ &\quad = \sum_{\substack{B\subseteq \mathbf{X}_{K\cup L}\\ B^{\downarrow K\cap L}=A^{\downarrow K\cap L}\\ B=B^{\downarrow K}\otimes B^{\downarrow L}}} \phi(B^{\downarrow K}) \cdot \psi(B^{\downarrow L})\\ &\quad = \sum_{\substack{D\subseteq \mathbf{X}_K\\ D^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} \sum_{\substack{C\subseteq \mathbf{X}_L\\ C^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} \phi(D)\cdot\psi(C)\\ &\quad = \left(\sum_{\substack{D\subseteq \mathbf{X}_K\\ D^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} \phi(D)\right)\cdot\left(\sum_{\substack{C\subseteq \mathbf{X}_L\\ C^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} \psi(C)\right).\end{aligned}$$

To finish the proof it is enough to substitute the corresponding expressions computed above into the formula from Definition 2, which is in this context in the form

$$m^{\downarrow K\cup L}(A)\cdot m^{\downarrow K\cap L}(A^{\downarrow K\cap L}) = m^{\downarrow K}(A^{\downarrow K})\cdot m^{\downarrow L}(A^{\downarrow L}).$$

Doing this we get

$$\begin{aligned}&m^{\downarrow K\cup L}(A)\cdot m^{\downarrow K\cap L}(A^{\downarrow K\cap L})\\ &\quad = \phi(A^{\downarrow K})\cdot\psi(A^{\downarrow L})\cdot\left(\sum_{\substack{D\subseteq \mathbf{X}_K\\ D^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} \phi(D)\right)\\ &\qquad\qquad \cdot\left(\sum_{\substack{C\subseteq \mathbf{X}_L\\ C^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} \psi(C)\right),\end{aligned}$$

$$\begin{aligned}&m^{\downarrow K}(A^{\downarrow K})\cdot m^{\downarrow L}(A^{\downarrow L})\\ &\quad = \left(\phi(A^{\downarrow K})\cdot\sum_{\substack{C\subseteq \mathbf{X}_L\\ C^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} \psi(C)\right)\\ &\qquad\qquad \cdot\left(\psi(A^{\downarrow L})\cdot\sum_{\substack{D\subseteq \mathbf{X}_K\\ D^{\downarrow K\cap L}=A^{\downarrow K\cap L}}} \phi(D)\right).\end{aligned}$$

□

*Corollary 3:* Let $m_1, m_2, m_3, \ldots$ be a sequence of basic assignments each of them factorizing with respect to $(K, L)$. If this sequence is convergent then also the limit basic assignment $\lim_{j\to+\infty} m_j$ factorizes with respect to $(K, L)$.

*Proof:* Since the considered frame of discernment $\mathbf{X}_N$ is finite, it is obvious that convergence of $m_1, m_2, \ldots$ implies also the convergence of all its marginals, i.e. also $\lim_{j\to+\infty} m_j^{\downarrow K}$, $\lim_{j\to+\infty} m_j^{\downarrow L}$ and $\lim_{j\to+\infty} m_j^{\downarrow K\cap L}$. The assumption of factorization of all $m_j$ says that, due to Theorem,

$$m_j^{\downarrow K\cup L}\cdot m_j^{\downarrow K\cap L} = m_j^{\downarrow K}\cdot m_j^{\downarrow L},$$

and therefore also

$$\lim_{j\to+\infty} m_j^{\downarrow K\cup L}\cdot \lim_{j\to+\infty} m_j^{\downarrow K\cap L} = \lim_{j\to+\infty} m_j^{\downarrow K}\cdot \lim_{j\to+\infty} m_j^{\downarrow L}.$$

□

From the point of view of possible applications it is important to realize that the notion of simple factorization introduced in Definition 3 is a direct analogy to the probabilistic simple factorization. Though we do not know whether one can introduce a meaningful analogy to the multiple factorization within Dempster-Shafer theory of evidence, the analogies to recursive factorization and decomposition are straightforward.

*Definition 4 (Recursive factorization):* Consider a finite sequence of nonempty sets $K_1, K_2, \ldots, K_r \subseteq N$. We say that basic assignment $m$ *factorizes with respect to* $(K_1, K_2, \ldots, K_r)$ if for each $i = 2, \ldots, r$ $m$ (simply) factorizes with respect to the pair $(K_1 \cup \ldots \cup K_{i-1}, K_i)$.

*Definition 5 (Decomposability):* Consider a finite sequence of nonempty sets $K_1, K_2, \ldots, K_r \subseteq N$ meeting RIP. We say that basic assignment $m$ is *decomposable with respect to* $(K_1, K_2, \ldots, K_r)$ if it recursively factorizes with respect to this sequence.

## V. Space-saving power of factorization

To persuade the reader that the notion of factorization is not interesting only from the theoretical point of view but that it is important also for applications, let us illustrate the space-saving power of factorization (decomposability) for very simple examples. As it can be seen even from the described trivial situations, the more complex models are taken into consideration the greater part of the storage demands are saved.

In Table I we consider four binary and two ternary models. The first line (*frame of discernment*) bears the information about the dimensionality of the considered basic assignments. Thus, the fourth line (*# of nonempty subsets*) shows how many numbers define a general basic assignment for the respective space of discernment (the maximum number of focal elements of a general basic assignment). The considered model (the type of factorization/decomposability) is described in the second line (*factorization*). Thus, for example, $\{1\}, \{2\}$ says that a basic assignment simply factorizing with respect to $(\{1\}, \{2\})$ is

Table I
SURVEY OF SPACE-SAVING POWER

| | Binary: $\mathbf{X}_i = \{a_i, \bar{a}_i\}$ | | | | Ternary: $\mathbf{X}_i = \{a_i, \bar{a}_i, \mathring{a}_i\}$ | |
|---|---|---|---|---|---|---|
| frame of discernment | $\mathbf{X}_{\{1,2\}}$ | $\mathbf{X}_{\{1,2,3\}}$ | $\mathbf{X}_{\{1,2,3,4\}}$ | $\mathbf{X}_{\{1,2,3,4\}}$ | $\mathbf{X}_{\{1,2\}}$ | $\mathbf{X}_{\{1,2,3\}}$ |
| factorization | $\{1\},\{2\}$ | $\{1,2\},\{2,3\}$ | $\{1,2,3\},\{2,3,4\}$ | $\{1,2\},\{2,3\},\{3,4\}$ | $\{1\},\{2\}$ | $\{1,2\},\{2,3\}$ |
| cardinality of the frame | 4 | 8 | 16 | 16 | 9 | 27 |
| # of nonempty subsets | 15 | 255 | 65 535 | 65 535 | 511 | 134 217 727 |
| $\lvert\{A : A = A^{\downarrow\{..\}} \otimes A^{\downarrow\{..\}}\rvert$ | 10 | 100 | 10 000 | 658 | 50 | 125 000 |
| # of subsets of each marginal | 4 | 16 | 256 | 16 | 8 | 512 |
| number of factors | $2 \times 3 = 6$ | $2 \times 15 = 30$ | $2 \times 255 = 510$ | $3 \times 15 = 45$ | $2 \times 7 = 14$ | $2 \times 511 = 1\ 022$ |
| space requirements | 0.4 | 0.118 | 0.0078 | 0.00069 | 0.027 | 0.0000076 |

considered in the first column. Similarly, $\{1,2\},\{2,3\},\{3,4\}$ means that the respective basic assignment recursively factorizes (is decomposable) with respect to $(\{1,2\},\{2,3\},\{3,4\})$. The fifth line ($\lvert\{A : A = A^{\downarrow\{..\}} \otimes A^{\downarrow\{..\}}\rvert$) shows how many subsets of the respective space of discernment meet the indicated property. For the last binary case (the fourth column) it means for how many subsets $A$ of $\mathbf{X}_{\{1,2,3,4\}}$

$$A = A^{\downarrow\{1,2\}} \otimes A^{\downarrow\{2,3\}} \otimes A^{\downarrow\{3,4\}}.$$

In any case the fifth line contains the maximum number of focal elements of a basic assignment factorizing with respect to the considered model, whereas the last but one line (*number of factors*) says the number of parameters by which the model is actually defined. The last line is a ratio

$$\textit{space requirements} = \frac{\textit{number of factors}}{\textit{\# of nonempty subsets}}$$

that is inversely proportional to the efficiency of the model: the lower the ration the more efficient the respective model is.

## VI. CONCLUSIONS

Inspired by factorization in probability theory, we have introduced an analogous notion in Dempster-Shafer theory of evidence. We have shown that it meets the basic property of probabilistic factorization that is anchored in the assertion widely known as *Factorization lemma*. We fully agree with the anonymous reviewer who stated: “The idea of generalizing the fundamental concepts from probability theory to belief functions is very natural.” It is simple and natural, and also the Dempster-Shafer version of Factorization lemma (presented here in a form of Theorem) is quite natural. This means, however, that conditional independence in Dempster-Shafer theory of evidence is to be defined as introduced here in Definition 2. Recall that the presented result is not the only one showing evidence in favor of this definition. As it was already presented in [2], Studený showed that the concept of conditional independence based on an application of conjunctive combination rule is not consistent with marginalization. He found two consistent basic assignments for which there does not exist a common extension manifesting the respective conditional independence (for more details and Studený example see [2]).

## REFERENCES


[1] B. Ben Yaghlane, Ph. Smets, and K. Mellouli, “Belief Function Independence: I. The Marginal Case,” *Int. J. of Approximate Reasoning*, vol. 29, no. 1, pp. 47–70, 2002.

[2] B. Ben Yaghlane, Ph. Smets, and K. Mellouli, “Belief Function Independence: II. The Conditional Case,” *Int. J. of Approximate Reasoning*, vol. 31, no. (1-2), pp. 31–75, 2002.

[3] I. Couso, S. Moral and P. Walley, “Examples of independence for imprecise probabilities,” in *Proceedings of ISIPTA’99*, G. de Cooman, F. G. Cozman, S. Moral, P. Walley, Eds., location and date, 1999, pp. 121–130.

[4] A. Dempster, “Upper and lower probabilities induced by a multi-valued mapping,” *Annals of Mathematical Statistics* vol. 38, pp. 325–339, 1967.

[5] R. Jiroušek, “On a conditional irrelevance relation for belief functions based on the operator of composition,” in *Dynamics of Knowledge and Belief, Proceedings of the Workshop at the 30th Annual German Conference on Artificial Intelligence,* Ch. Beierle, G. Kern-Isberner, Eds., Fern Universität in Hagen, Osnabrück, 2007, pp. 28-41.

[6] R. Jiroušek and J. Vejnarová, “Compositional models and conditional independence in evidence theory.” Accepted to *Int. J. Approx. Reasoning*.

[7] G. J. Klir, *Uncertainty and Information. Foundations of Generalized Information Theory.* Wiley, Hoboken, 2006.

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

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

[10] P. P. Shenoy, “Conditional independence in valuation-based systems,” *Int. J. of Approximate Reasoning*, vol. 10, no. 3, pp. 203–234, 1994.

[11] P. P. Shenoy, “Binary join trees for computing marginals in the Shenoy-Shafer architecture,” *Int. J. of Approximate Reasoning*, vol. 17, no. (2-3), pp. 239–263, 1997.

[12] M. Studený, “Formal properties of conditional independence in different calculi of AI,” in *Proceedings of European Conference on Symbolic and quantitative Approaches to Reasoning and Uncertainty ECSQARU’93*, K. Clarke, R. Kruse, S. Moral, Eds., location and date, Springer-Verlag, 1993, pp. 341–351.

[13] M. Studený, “On stochastic conditional independence: the problems of characterization and description,” *Annals of Mathematics and Artificial Intelligence,* vol. 35, p. 323-341, 2002.