A note on simple factorization in Dempster-Shafer theory of evidence

Radim Jiroušek

In honour of Petr Hájek
on the occasion of his 70th birthday

ABSTRACT. The paper introduces (in our knowledge first) attempt to define the concept of factorization within the Dempster-Shafer theory of evidence. In the same way as in probability theory, the presented 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.

1 Introduction

It was at the beginning of 80's of the last century when I first heard about Dempster-Shafer theory of Evidence, and it was Petr Hájek who was lecturing on this topic. In fact it was him who organized (and still is organizing) a series of working seminars (so called Hájek's seminars) at the Institute of Mathematics of the Czechoslovak Academy of Sciences, where not only me but many regular attendees from several research and university institutions heard about some interesting topics first time.

To tell the truth, in those days, when mainframe computers with several hundreds of kilobytes of memory represented the best available computational technique, I considered Dempster-Shafer theory to be a rather academic topic. At that time I could not imagine that a model with super-exponential space complexity could ever be applied to problems of practice. And yet, in the second half of 80's, Petr Hájek became an enthusiastic booster of this technique [Háj87, Háj92, HH92b, HH92a, Háj93, Háj94]. And as it has appeared later he was right. Since that time the computational tools have empowered in such a way that these models have become computationally tractable.

Naturally, before Dempster-Shafer models started being employed to problems of practice, there was a great flowering of probabilistic (and also possibilistic) models with their exponential computational complexity. This was because a substantial decrease of computational complexity was achieved with the help of models taking advantage of the concept of conditional independence. So, representing multidimensional probability distributions in a form of Graphical Markov Models (GMMs) (e.g. Bayesian networks) made it possible to store in a computer memory distributions of hundreds (or even thousands) of dimensions. However, studying properly probabilistic GMMs 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 the notion of *factorization*, which in probability theory (due to factorization lemma presented here as Lemma 1) coincide with conditional independence. Going into details, one can notice that the notion of factorization has been introduced in several different ways in probability theory, and some others can still be studied.

This is why in this paper we shall first briefly analyze the notion of factorization in probability theory and only afterwards will generalize the simplest definition for the Dempster-Shafer theory of evidence.

2 Probabilistic Factorization

In this section we 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 π (or ν) 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

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

Let us remark that for $K = \emptyset$ we get $\pi^{\downarrow \emptyset} = 1$.

An 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 π and three disjoint groups of variables $X_K = \{X_i\}_{i\in K}$, $X_L = \{X_i\}_{i\in L}$ and $X_M = \{X_i\}_{i\in M}$ ($K, L, M \subset N$, $K \neq \emptyset \neq L$) having their values in $\mathbf{X}_K$, $\mathbf{X}_L$ and $\mathbf{X}_M$, respectively. We say that X_K and X_L are *conditionally independent given* X_M (with respect to probability measure π) if for all $x \in \mathbf{X}_{K\cup L\cup M}$

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

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*[1]) *independent*, which is usually denoted by a simplified notation: $K \perp\!\!\!\perp L\ [\pi]$.

As already mentioned in Introduction, 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 be either found in [Lau96] or can be directly deduced as trivial consequences of properties presented there.

Simple factorization

Consider two nonempty sets $K, L \subseteq N$. We say that π *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[2]

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

holds true.

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

COROLLARY 2 *π 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 3 *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 $\lim\limits_{j\to+\infty} \pi_j$ factorizes with respect to (K, L).*

[1]Some author call it marginal independence.

[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 $\{x\}^{\downarrow K}$.

Multiple factorization

Consider a finite system of nonempty subsets $K_1, K_2, \ldots, K_p$ of a set N. We say that π *factorizes with respect to* $(K_1, K_2, \ldots, K_p)$ if there exist p functions $(i = 1, 2, \ldots, p)$

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

such that for all $x \in \mathbf{X}_{K_1 \cup \ldots \cup K_p}$

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

REMARK 4 *In this general case one can (using Lemma 1) derive a system of conditional independence relations valid for a measure π factorizing with respect to $(K_1, K_2, \ldots, K_p)$ but no assertion that could be considered a direct analogy to any of the preceding Corollaries hold true. This is why the following type of factorization is often considered.*

Recursive factorization

Consider a finite system of nonempty subsets $K_1, K_2, \ldots, K_p$ of a set N. We say that π *recursively factorizes with respect to* $(K_1, K_2, \ldots, K_p)$ if for each $i = 2, \ldots, p$ π (simply) factorizes with respect to the pair

$$(K_1 \cup \ldots \cup K_{i-1}, K_i).$$

REMARK 5 *Using Corollary 2 iteratively one can get a formula expressing the multidimensional measure $\pi^{\downarrow K_1 \cup \ldots \cup K_p}$ with the help of its respective marginals*[3]

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

So we are getting a trivial assertion saying that if π recursively factorizes with respect to $K_1, K_2, \ldots, K_p$ 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 3 holds true.

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

Marginal factorization

Marginal factorization, which is introduced in this paragraph is usually not considered by other authors. It is weaker than the recursive factorization (and in a sense stronger than factorization).

Consider a finite system of nonempty subsets $K_1, K_2, \ldots, K_p$ of a set N. We say that π *marginally factorizes with respect to* $(K_1, K_2, \ldots, K_p)$ if $\pi^{\downarrow K_1 \cup \ldots \cup K_p}$ is uniquely given by its marginals $\pi^{\downarrow K_1}, \pi^{\downarrow K_2}, \ldots, \pi^{\downarrow K_p}$. More precisely, for this type of factorization we assume that there exists a function $\mathcal{F}$ such that

$$\pi^{\downarrow K_1 \cup \ldots \cup K_p} = \mathcal{F}(\pi^{\downarrow K_1}, \pi^{\downarrow K_2}, \ldots, \pi^{\downarrow K_p}).$$

As an interesting example of this type of factorization may serve the following function[4] $\mathcal{F}$

$$\mathcal{F}(\pi^{\downarrow K_1}, \pi^{\downarrow K_2}, \ldots, \pi^{\downarrow K_p}) = \arg\max\{\mathbf{H}(\nu)\},$$

where the maximization is performed over all measures ν on $\mathbf{X}_{K_1 \cup \ldots \cup K_p}$ for which $\nu^{\downarrow K_1} = \pi^{\downarrow K_1}, \nu^{\downarrow K_2} = \pi^{\downarrow K_2}, \ldots, \nu^{\downarrow K_p} = \pi^{\downarrow K_p}$.

REMARK 6 *It is obvious that this type of factorization strongly depends on the function $\mathcal{F}$. Generally, no factorization lemma holds true for marginal factorization. However, if one considers a "reasonable" function $\mathcal{F}$ (for example if it is continuous in a sense) then Corollary 3 holds true.*

Decomposition

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

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

LEMMA 7 (Decomposition lemma) *If $(K_1, K_2, \ldots, K_p)$ meets RIP, then π factorizes with respect to $(K_1, K_2, \ldots, K_p)$ if and only if it recursively factorizes with respect to $(K_1, K_2, \ldots, K_p)$.*

In the literature, measures factorizing with respect to systems of sets meeting (after a possible reordering) RIP are usually called *decomposable measures*.

[4] $\mathbf{H}$ denotes the classical Shannon entropy

$$\mathbf{H}(\nu) = -\sum_{x:\nu(x)>0} \nu(x)\log(\nu(x)).$$

REMARK 8 *It is interesting to notice that all these definitions coincide if one considers only two sets of indices: π factorizes with respect to (K, L) if and only if it marginally factorizes with respect to (K, L). Notice also that a two-element sequence meets always RIP. This is why not many authors distinguish different types of factorization.*

3 D-S 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 important to keep this in mind, 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 assignment 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

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

4 Independence and factorization in D-S Theory

Let us now present a generally accepted notion of independence ([YSM02a, She94, Stu93]).

DEFINITION 9 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*[5] 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 [YSM02b, CMW99, Kli06, She94, Stu93]). In this text we will use the generalization, which was introduced in [JV] and which differs from the notion used in [YSM02b, She94, Stu93]).

DEFINITION 10 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 all $A \subseteq \mathbf{X}_{K \cup L \cup M}$:

- if $A = A^{\downarrow K \cup M} \otimes A^{\downarrow L \cup M}$ then

$$m^{\downarrow K \cup L \cup M}(A) \cdot m^{\downarrow M}(A^{\downarrow M}) = m^{\downarrow K \cup M}(A^{\downarrow K \cup M}) \cdot m^{\downarrow L \cup M}(A^{\downarrow L \cup M}),$$

- if $A \neq A^{\downarrow K \cup M} \otimes A^{\downarrow L \cup M}$ then $m^{\downarrow K \cup L \cup M}(A) = 0$.

[5] Couso et al. [CMW99] call this independence *independence in random sets*, Klir [Kli06] *non-interactivity*.

Notice that for $M = \emptyset$ the concept coincides with Definition 9. In addition to this, and this is even more important, it was proven in [JV] that this notion meets all the properties required from the notion of conditional independence, so-called *semigraphoid properties* ([Lau96, Stu93]):

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

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

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

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

To be honest, we have to recall that all these properties (both the semigraphoid properties and the fact that the notion is a generalization of the unconditional independence) hold true also for the concept of conditional independence used, for example, by Shenoy [She94] and Studený [Stu93] (which is the same as the *conditional non-interactivity* used by Ben Yaghlane *et al.* in [YSM02b]). In spite of the fact that their term is used also by several other authors, we do not expect that it satisfies a characterization property similar to the one proven below in Factorization lemma.

DEFINITION 11 (Simple factorization) Consider two nonempty sets $K, L \subseteq N$. We say that basic assignment *m factorizes with respect to* (K, L) if it satisfy the following two properties:

- for all $A \subseteq \mathbf{X}_{K\cup L}$, for which $A \neq A^{\downarrow K} \otimes A^{\downarrow L}$, $m(A) = 0$;
- there exist two nonnegative set functions
$$\phi : \mathcal{P}(\mathbf{X}_K) \longrightarrow [0, +\infty) \quad \text{and} \quad \psi : \mathcal{P}(\mathbf{X}_L) \longrightarrow [0, +\infty),$$
such that for each $A \subseteq \mathbf{X}_{K\cup L}$, for which $A = A^{\downarrow K} \otimes A^{\downarrow L}$, we have
$$m^{\downarrow K\cup L}(A) = \phi(A^{\downarrow K}) \cdot \psi(A^{\downarrow L})$$

LEMMA 12 (Factorization lemma) *Let $K, L \subseteq N$ be nonempty. m factorizes with respect to (K, L) if and only if*

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

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 $K \setminus L \perp\!\!\!\perp L \setminus K \,|\, K \cap L\ [m]$. So, to prove implication

$$K \setminus L \perp\!\!\!\perp L \setminus K \,|\, K \cap L\ [m] \quad \Longrightarrow \quad m \text{ factorizes with respect to } (K, L)$$

is trivial. It is enough to take

$$\phi = m^{\downarrow L} \qquad\qquad \psi = \frac{m^{\downarrow L}}{m^{\downarrow K\cap L}},$$

(for $m^{\downarrow K\cap L}(x^{\downarrow K\cap L}) = 0$ take $\frac{0}{0} = 0$).

To prove the opposite implication consider two functions ϕ and ψ meeting the properties required by Definition 11, 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

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

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 summarize only over 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 11, $m(B) = 0$. 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

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

Using this we get

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

To finish the proof it is enough to substitute into the formula from Definition 10, which is in this context in the form

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

the corresponding expressions computed above:

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

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

$\square$

COROLLARY 13 *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 Lemma 12,

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

and therefore also

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

□

REMARK 14 *The notion introduced in Definition 11 is an analogy to the probabilistic simple factorization. Therefore, one can directly introduce also recursive factorization and decomposition for Dempster-Shafer theory of evidence. The problem whether analogical notions to multiple factorization and marginal factorization are meaningful remains at the moment open.*

5 Application of factorization to model construction

As said in Introduction, factorization is fully employed in probabilistic GMMs. So it is quite natural, that some authors generalized the ideas of GMMs and started considering analogous models within the framework of Dempster-Shafer theory. In this paper, we shall go different way. We shall describe a generalization of an alternative probabilistic approach, which is based on the application of the operator of composition. After describing this operator, we shall briefly illustrate its application to the construction of *compositional models* in the framework of Dempster-Shafer.

Operator of Composition

Let K and L be two subsets of N. At this moment we do not pose any restrictions on K and L; they may be but need not be disjoint, one may be a subset of the other. We even admit that one or both of them are empty but this is just a theoretical possibility without any practical impact[6]. Let m_1 and m_2 be basic assignments on $\mathbf{X}_K$ and $\mathbf{X}_L$, respectively.

[6]Notice that basic assignment m on $\mathbf{X}_\emptyset$ is defined $m(\emptyset) = 1$. Let us note that this is the only case when we accept $m(\emptyset) > 0$, otherwise $m(\emptyset) = 0$ according to the classical definitions of basic assignment and belief function, see [Sha76].

DEFINITION 15 For two arbitrary basic assignments m_1 on $\mathbf{X}_K$ and m_2 on $\mathbf{X}_L$ a *composition* $m_1 \triangleright m_2$ is defined for all $C \subseteq \mathbf{X}_{K\cup L}$ by one of the following expressions:

[**a**] if $m_2^{\downarrow K\cap L}(C^{\downarrow K\cap L}) > 0$ and $C = C^{\downarrow K} \otimes C^{\downarrow L}$ then

$$(m_1 \triangleright m_2)(C) = \frac{m_1(C^{\downarrow K}) \cdot m_2(C^{\downarrow L})}{m_2^{\downarrow K\cap L}(C^{\downarrow K\cap L})};$$

[**b**] if $m_2^{\downarrow K\cap L}(C^{\downarrow K\cap L}) = 0$ and $C = C^{\downarrow K} \times \mathbf{X}_{L\setminus K}$ then

$$(m_1 \triangleright m_2)(C) = m_1(C^{\downarrow K});$$

[**c**] in all other cases

$$(m_1 \triangleright m_2)(C) = 0.$$

REMARK 16 *Notice what this definition yields in the following trivial situations:*

- *if* $K \supseteq L$ *then* $m_1 \triangleright m_2 = m_1$ *(therefore, the operator of composition is idempotent);*
- *if* $K \cap L = \emptyset$ *then for each* $C \subseteq \mathbf{X}_{K\cup L}$

$$m_1 \triangleright m_2(C) = \begin{cases} m_1(C^{\downarrow K}) \cdot m_2(C^{\downarrow L}) & \textit{for } C = C^{\downarrow K} \times C^{\downarrow L}, \\ 0 & \textit{otherwise}, \end{cases}$$

(i.e. $m_1 \triangleright m_2$ *is a basic assignment of independent groups of variables* X_K *and* X_L*).*

REMARK 17 *Before presenting basic properties of this operator, we stress that*

- *operator* $\triangleright$ *is different from the famous Dempster's rule of combination*[7] *(these two rules coincide only under very special conditions, for details see [JV09]);*
- *it is neither commutative nor associative.*

[7] Recall that, for example, in contrast to Dempster's rule of combination, the operator of composition is idempotent.

Table 1. Basic assignments

A	$m_1(A)$	A	$m_2(A)$	A	$m_3(A)$
$\{a_1a_2\}$	0.5	$\{a_3\}$	0.5	$\{a_2a_3a_4\}$	0.5
$\mathbf{X}_1 \times \{\bar{a}_2\}$	0.5	$\{\bar{a}_3\}$	0.5	$\{\bar{a}_2\bar{a}_3\} \times \mathbf{X}_4$	0.5

The following two assertions were proven in [JVD07]

LEMMA 18 *For arbitrary two basic assignments* m_1 *on* $\mathbf{X}_K$ *and* m_2 *on* $\mathbf{X}_L$ *the following properties hold true:*

1. $m_1 \triangleright m_2$ *is a basic assignment on* $\mathbf{X}_{K\cup L}$*;*
2. $(m_1 \triangleright m_2)^{\downarrow K} = m_1$*;*
3. $m_1 \triangleright m_2 = m_2 \triangleright m_1 \iff m_1^{\downarrow K\cap L} = m_2^{\downarrow K\cap L}$.

LEMMA 19 *For arbitrary basic assignment* m *on* $\mathbf{X}_M$ $(M = K \cup L)$ *the following properties hold true:*

1. $m^{\downarrow K} \triangleright m = m$*;*
2. $m = m^{\downarrow K} \triangleright m^{\downarrow L} \iff K \setminus L \perp\!\!\!\perp L \setminus K \mid K \cap L\,[m]$.

Thus, having a finite system of (low-dimensional) basic assignment m_1, $m_2, \ldots, m_p$ one can consider a (multidimensional) basic assignment

$$m_1 \triangleright m_2 \triangleright m_3 \triangleright \ldots \triangleright m_p = (\ldots((m_1 \triangleright m_2) \triangleright m_3) \triangleright \ldots \triangleright m_p).$$

(Since the operator is not associative we have to stress that the operator is applied subsequently from right to left.) Such a basic assignment is called a *compositional model*.

Example

Let for $i = 1, 2, 3, 4$, $\mathbf{X}_i = \{a_i, \bar{a}_i\}$ and consider three basic assignments m_1, m_2 and m_3 given in Table 1. Here we present only focal elements of the respective basic assignments. This means that basic assignments equal 0 for all sets which are not presented in the tables. From the table we can also see that m_1, m_2 and m_3 are basic assignments on $\mathbf{X}_1 \times \mathbf{X}_2$, $\mathbf{X}_3$ and $\mathbf{X}_2 \times \mathbf{X}_3 \times \mathbf{X}_4$, respectively. First, notice that the given basic assignments are pairwise projective, i.e. $m_1^{\downarrow\{2\}} = m_3^{\downarrow\{2\}}$ and $m_2 = m_3^{\downarrow\{3\}}$ (m_1 and m_2 are trivially projective because they are defined on disjoint spaces). This is because both $m_1^{\downarrow\{2\}}$ and $m_3^{\downarrow\{2\}}$ acquire the same values for all (in our case only two) focal elements:$\{a_2\}$ and $\{\bar{a}_2\}$

$$m_1^{\downarrow\{2\}}(\{a_2\}) = m_3^{\downarrow\{2\}}(\{a_2\}) = 0.5,$$
$$m_1^{\downarrow\{2\}}(\{\bar{a}_2\}) = m_3^{\downarrow\{2\}}(\{\bar{a}_2\}) = 0.5,$$

and similarly

$$m_2(\{a_3\}) = m_3^{\downarrow\{3\}}(\{a_3\}) = 0.5,$$
$$m_2(\{\bar{a}_3\}) = m_3^{\downarrow\{3\}}(\{\bar{a}_3\}) = 0.5.$$

Now, we shall show that, in spite of the mentioned projectiveness (and due to the already mentioned non-associativity of the operator of composition), one can construct two different models corresponding to permutations m_1, m_2, m_3 and m_2, m_3, m_1.

First, compute $m_1 \triangleright m_2$. Since it is a composition of two basic assignments defined on disjoint spaces, the number of focal elements of $m_1 \triangleright m_2$ equals a product of numbers of focal elements of assignments m_1 and m_2: $2 \times 2 = 4$. The respective values of the composed basic assignments are computed according to case [**a**] of Definition 15:

$$(m_1 \triangleright m_2)(\{a_1 a_2 a_3\}) = \frac{m_1(\{a_1 a_2\}) \cdot m_2(\{a_3\})}{m_2^{\downarrow\emptyset}(\emptyset)} = \frac{0.5 \cdot 0.5}{1} = 0.25,$$

$$(m_1 \triangleright m_2)(\{a_1 a_2 \bar{a}_3\}) = \frac{m_1(\{a_1 a_2\}) \cdot m_2(\{\bar{a}_3\})}{m_2^{\downarrow\emptyset}(\emptyset)} = \frac{0.5 \cdot 0.5}{1} = 0.25,$$

$$(m_1 \triangleright m_2)(\mathbf{X}_1 \times \{\bar{a}_2 a_3\}) = \frac{m_1(\mathbf{X}_1 \times \{\bar{a}_2\}) \cdot m_2(\{a_3\})}{m_2^{\downarrow\emptyset}(\emptyset)} = \frac{0.5 \cdot 0.5}{1} = 0.25,$$

$$(m_1 \triangleright m_2)(\mathbf{X}_1 \times \{\bar{a}_2 \bar{a}_3\}) = \frac{m_1(\mathbf{X}_1 \times \{\bar{a}_2\}) \cdot m_2(\{\bar{a}_3\})}{m_2^{\downarrow\emptyset}(\emptyset)} = \frac{0.5 \cdot 0.5}{1} = 0.25.$$

From this one immediately sees that also its marginal basic assignment $(m_1 \triangleright m_2)^{\downarrow\{2,3\}}$ has four focal elements ($\{a_2 a_3\}$, $\{a_2 \bar{a}_3\}$, $\{\bar{a}_2 a_3\}$, $\{\bar{a}_2 \bar{a}_3\}$) and therefore $(m_1 \triangleright m_2)^{\downarrow\{2,3\}}$ and m_3 cannot be projective. Therefore, computation of $m_1 \triangleright m_2 \triangleright m_3$ will be a little bit more complicated. Case [**a**] of Definition 15 applies to two focal elements:

$$(m_1 \triangleright m_2 \triangleright m_3)(\{a_1a_2a_3a_4\}) = \frac{(m_1 \triangleright m_2)(\{a_1a_2a_3\}) \cdot m_3(\{a_2a_3a_4\})}{m_3^{\downarrow\{2,3\}}(\{a_2a_3\})} = \frac{0.25 \cdot 0.5}{0.5} = 0.25,$$

$$(m_1 \triangleright m_2 \triangleright m_3)(\mathbf{X}_1 \times \{\bar{a}_2\bar{a}_3\} \times \mathbf{X}_4) = \frac{(m_1 \triangleright m_2)(\mathbf{X}_1 \times \{\bar{a}_2\bar{a}_3\}) \cdot m_3(\{\bar{a}_2\bar{a}_3\} \times \mathbf{X}_4)}{m_3^{\downarrow\{2,3\}}(\{\bar{a}_2\bar{a}_3\})} = \frac{0.25 \cdot 0.5}{0.5} = 0.25.$$

Values of the remaining two focal elements are assigned according to case [**b**] of Definition 15:

$$(m_1 \triangleright m_2 \triangleright m_3)(\{a_1a_2\bar{a}_3\} \times \mathbf{X}_4) = (m_1 \triangleright m_2)(\{a_1a_2\bar{a}_3\}) = 0.25,$$
$$(m_1 \triangleright m_2 \triangleright m_3)(\mathbf{X}_1 \times \{\bar{a}_2a_3\} \times \mathbf{X}_4) = (m_1 \triangleright m_2)(\mathbf{X}_1 \times \{\bar{a}_2a_3\}) = 0.25.$$

Computation of $m_2 \triangleright m_3 \triangleright m_1$ is simple. Since m_2 is marginal to m_3 it follows from property (1) of Lemma 19 that $m_2 \triangleright m_3 = m_3$. The remaining computation of $m_3 \triangleright m_1$ consists in application of the formula from case [**a**] of Definition 15 just to two focal elements $\{a_1a_2a_3a_4\}$ and $\mathbf{X}_1 \times \{\bar{a}_2\bar{a}_3\} \times \mathbf{X}_4$:

$$(m_3 \triangleright m_1)(\{a_1a_2a_3a_4\}) = \frac{m_3(\{a_2a_3a_4\}) \cdot m_1(\{a_1a_2\})}{m_1^{\downarrow\{2\}}(\{a_2\})} = \frac{0.5 \cdot 0.5}{0.5} = 0.5,$$

$$(m_3 \triangleright m_1)(\mathbf{X}_1 \times \{\bar{a}_2\bar{a}_3\} \times \mathbf{X}_4) = \frac{m_3(\{\bar{a}_2\bar{a}_3\} \times \mathbf{X}_4) \cdot m_1(\mathbf{X}_1 \times \{\bar{a}_2\})}{m_2^{\downarrow\{2\}}(\{\bar{a}_2\})} = \frac{0.5 \cdot 0.5}{0.5} = 0.5.$$

Table 2. Composed basic assignments.

C	$(m_1 \triangleright m_2 \triangleright m_3)(C)$	$(m_2 \triangleright m_3 \triangleright m_1)(C)$
$\{a_1a_2a_3a_4\}$	0.25	0.5
$\{a_1a_2\bar{a}_3\} \times \mathbf{X}_4$	0.25	0
$\mathbf{X}_1 \times \{\bar{a}_2a_3\} \times \mathbf{X}_4$	0.25	0
$\mathbf{X}_1 \times \{\bar{a}_2\bar{a}_3\} \times \mathbf{X}_4$	0.25	0.5

Both the resulting 4-dimensional basic assignments are recorded in Table 2 (recall once more that for all the other subsets of $\mathbf{X}_1 \times \mathbf{X}_2 \times \mathbf{X}_3 \times \mathbf{X}_4$, different from those included in Table 2, both the assignments equal 0).

6 Conclusions

Inspired by the (simple) 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*. Existence of its Dempster-Shafer version (presented here as Lemma 12) forms a new evidence supporting the definition of conditional independence introduced first in [Jir07][8] and studied in more details in [JV] (see Definition 10). The last section was included 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, for multidimensional model construction.

Acknowledgement

The research was partially supported by Ministry of Education of the Czech Republic under grants no 1M0572 and 2C06019, and by Czech Science Foundation under the grants no. ICC/08/E010 and 201/09/1891.

BIBLIOGRAPHY

[YSM02a] B. Ben Yaghlane, P. Smets, and K. Mellouli. Belief function independence: I. The marginal case. *Int. J. Approximate Reasoning*, 29:47–70, 2002.

[YSM02b] B. Ben Yaghlane, P. Smets, and K. Mellouli. Belief function independence: II. The conditional case. *Int. J. Approximate Reasoning*, 31:31–75, 2002.

[CMW99] I. Couso, S. Moral, and P. Walley. Examples of independence for imprecise probabilities. In G. De Cooman, F. G. Cozman, S. Moral, and P. Walley, editors, *Proceedings of ISIPTA'99*, pages 121–130, 1999.

[Háj87] P. Hájek. Dempster-Shaferova teorie evidence a expertní systémy. In *Metody umělé inteligence a expertní systémy III*, pages 54–61, Prague, 1987. ČSVTS.

[Háj92] P. Hájek. Dempster-Shafer theory—what it is and how (not) to use it. In *SOFSEM '92*, pages 19–24, Brno, 1992. UVT MU.

[Háj93] P. Hájek. Deriving Dempster's rule. In B. Bouchon-Meunier, L. Valverde, and R. R. Yager, editors, *Uncertainty in Intelligent Systems*, pages 75–83, Amsterdam, 1993. North-Holland.

[Háj94] P. Hájek. Systems of conditional beliefs in Dempster-Shafer theory and expert systems. *Int. J. General Systems*, 22(2):113–124, 1994.

[HH92a] P. Hájek and D. Harmanec. An exercise in Dempster-Shafer theory. *Int. J. General Systems*, 20(2):137–142, 1992.

[HH92b] P. Hájek and D. Harmanec. On belief functions (the present state of Dempster-Shafer theory). In V. Mařík, editor, *Advanced in Artificial Intelligence*, pages 286–307, Berlin, 1992. Springer.

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

[Jir07] R. Jiroušek. On a conditional irrelevance relation for belief functions based on the operator of composition. In C. Beierle and G. Kern-Isberner, editors, *Dynamics of Knowledge and Belief. Proceedings of the Workshop at the 30th Annual German Conference on Artificial Intelligence*, pages 28–41, Osnabrück, 2007. Fern Universität in Hagen.

[JV] R. Jiroušek and J. Vejnarová. Compositional models and conditional independence in evidence theory. Submitted.

[JV09] R. Jiroušek and J. Vejnarová. There are combinations and compositions in Dempster-Shafer theory of evidence. In T. Kroupa and J. Vejnarová, editors, *Proceedings of the 8th Workshop on Uncertainty Processing—WUPES'09*, pages 100–111, 2009.

[JVD07] R. Jiroušek, J. Vejnarová, and M. Daniel. Compositional models for belief functions. In G. De Cooman, J. Vejnarová, and M. Zaffalon, editors, *Proceedings of 5th International Symposium on Imprecise Probability: Theories and Applications ISIPTA'07*, pages 243–252, Prague, 2007.

[Kli06] G. J. Klir. *Uncertainty and Information. Foundations of Generalized Information Theory.* Wiley, Hoboken, 2006.

[Lau96] S. L. Lauritzen. *Graphical Models.* Oxford University Press, 1996.

[Sha76] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, Princeton NJ, 1976.

[She94] P. P. Shenoy. Conditional independence in valuation-based systems. *Int. J. Approximate Reasoning*, 10:203–234, 1994.

[Stu93] M. Studený. Formal properties of conditional independence in different calculi of AI. In K. Clarke, R. Kruse, and S. Moral, editors, *Proceedings of European Conference on Symbolic and quantitative Approaches to Reasoning and Uncertainty ECSQARU'93*, pages 341–351. Springer, 1993.

Radim Jiroušek
Faculty of Management, University of Economics
Jarošovská 1117/II, CZ377 01 Jindřichův Hradec

and

Institute of Information Theory and Automation
Academy of Sciences of the Czech Republic
Pod Vodárenskou věží 4, CZ182 08 Prague
Czech Republic
Email: radim@utia.cas.cz