On Open Problems Connected with Application of the Iterative Proportional Fitting Procedure to Belief Functions

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

Václav Kratochvíl
Inst. of Inform. Theory and Automation
Acad. of Sciences, Prague, Czech Republic
v.kratochvil@gmail.com

Abstract

In probability theory, Iterative Proportional Fitting Procedure can be used for construction of a joint probability measure from a system of its marginals. The present paper studies a possibility of application of an analogous procedure for belief functions, which was made possible by the fact that there exist operators of composition for belief functions.

In fact, two different procedures based on two different composition operators are introduced. The procedure based on the composition derived from the Dempster's rule of combination is of very high computational complexity and, from the theoretical point of view, practically nothing is known about its behavior. The other one, which uses the composition derived from the notion of factorization, is much more computationally efficient, and its convergence is guaranteed by a theorem proved in this paper.

Keywords. Marginal problem, belief functions, algorithm, multidimensional model, convergence.

1 Introduction

In probability theory, by a marginal problem we understand a task to find out whether there exists a joint probability measure having a given system of low-dimensional measures for its marginals, and/or the problem how to find such a joint probability measure. In statistics this problem appears, for example, as a subtask of multidimensional contingency tables analysis. In 1980s, the problem was often solved in connection with a design of probabilistic knowledge-based systems [1, 10, 13]. In these expert systems, marginal measures represent pieces of local knowledge and the looked for multidimensional measure represents a knowledge base.

For a solution of a discrete marginal problem famous Iterative Proportional Fitting Procedure (IPFP) was suggested by Deming and Stephan in 1940 [3]. Though this iterative procedure was applied to practical problems since that time, it was only in 1975 when Csiszár proved its convergence [2].

The goal of this paper is to show that an analogous iterative procedure can be, in principle, applied also for construction of a multidimensional belief function. However, as the title of the paper suggests, this application is connected with several open questions.

1.1 Notation

In this paper we use the notation from the ISIPTA 2011 paper [4]: $\mathbb{X}_N = \mathbb{X}_1 \times \mathbb{X}_2 \times \ldots \times \mathbb{X}_n$, denotes a finite multidimensional space, and its subspaces (for all $K \subseteq N$) are denoted by

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

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

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

By a *join* of two sets $A \subseteq \mathbb{X}_K$ and $B \subseteq \mathbb{X}_L$ we understand a set

$$A \bowtie B = \{x \in \mathbb{X}_{K \cup L} : x^{\downarrow K} \in A \ \ \& \ \ x^{\downarrow L} \in B\}.$$

Let us note that if K and L are disjoint, then $A \bowtie B = A \times B$, if $K = L$ then $A \bowtie B = A \cap B$, and, generally, for $C \subseteq \mathbb{X}_{K \cup L}$, C is a subset of $C^{\downarrow K} \bowtie C^{\downarrow L}$, which may be proper.

A *basic assignment* m on $\mathbb{X}_K$ ($K \subseteq N$) is a real valued function on $\mathcal{P}(\mathbb{X}_K)$, for which

$$\sum_{\emptyset \neq A \subseteq \mathbb{X}_K} m(A) = 1.$$

Notice that in agreement with Shenoy's papers (see e.g., [12]) we admit also negative values of a basic assignment. This is why we will call a basic assignment to be *proper* if all its values are nonnegative. If

$m(A) \neq 0$, then A is said to be a *focal element* of m. Considering two proper basic assignments m_1, m_2 on the same space $\mathbb{X}_K$, we say that m_1 is *dominated* by m_2, if for all $A \subseteq \mathbb{X}_K$: $m_1(A) > 0 \Longrightarrow m_2(A) > 0$.

Having a basic assignment m on $\mathbb{X}_K$ one can consider its *marginal assignments*. On $\mathbb{X}_L$ (for $L \subseteq K$) it is defined (for each $\emptyset \neq B \subseteq \mathbb{X}_L$):

$$m^{\downarrow L}(B) = \sum_{A \subseteq \mathbb{X}_K : A^{\downarrow L} = B} m(A).$$

Each basic assignment m on $\mathbb{X}_K$ can uniquely be represented by its *commonality function*, which is a set function $Q : \mathcal{P}(\mathbb{X}_K) \longrightarrow [0, +\infty)$ defined for each $A \subseteq \mathbb{X}_K$

$$Q(A) = \sum_{A \subseteq B \subseteq \mathbb{X}_K} m(B).$$

Recall the formula from [11] yielding for each commonality function the respective basic assignment:

$$m(A) = \sum_{A \subseteq B \subseteq \mathbb{X}_K} (-1)^{|B \setminus A|} Q(B)$$

for each $A \subseteq \mathbb{X}_K$.

1.2 Operators of composition

In this paper we will take advantage of the fact that the probabilistic IPFP can easily (and elegantly) be expressed with the help of the so called operator of composition [5] that was defined in ISIPTA paper [8] also for belief functions. In [6] (see also an extended version of this conference contribution, which is to appear in IJAR [7]) it was shown that the operator of composition can also be defined within the Shenoy's valuation based systems (VBS) [12] that, as a generic uncertainty calculus, covers not only probability theory but also some other uncertainty calculi like Spohns epistemic belief theory, Dempster-Shafer belief function theory, and others.

In VBS's the operator of composition is derived from the operation of *combination* $\oplus$ and its inverse operation called *removal* $\ominus$. For two basic assignments m_1, m_2 on $\mathbb{X}_K$, $\mathbb{X}_L$, respectively, the operator of composition is defined as

$$m_1 \triangleright m_2 = m_1 \oplus m_2 \ominus m_2^{\downarrow K \cap L}, \tag{1}$$

from which one immediately sees its semantics: we combine knowledge contained in m_1 and m_2, and to prevent double counting of knowledge when double counting matters, we remove the knowledge contained in $m_2^{\downarrow K \cap L}$.

In Dempster-Shafer theory, the role of this general operator of composition $\oplus$ is played quite naturally by the Dempster's rule of combination $\oplus_D$. Thus, for m_1, m_2 on $\mathbb{X}_K$, $\mathbb{X}_L$, respectively, for each nonempty $A \subseteq \mathbb{X}_{K \cup L}$

$$(m_1 \oplus_D m_2)(A) = \Gamma^{-1} \sum_{B \subseteq \mathbb{X}_K, C \subseteq \mathbb{X}_L : B \bowtie C = A} m_1(B) \cdot m_2(C),$$

where Γ is the normalization factor

$$\Gamma = \sum_{B \subseteq \mathbb{X}_K, C \subseteq \mathbb{X}_L : B \bowtie C \neq \emptyset} m_1(B) \cdot m_2(C).$$

It is not an easy task to specify in terms of basic assignments the removal operator that should be an inverse to the Dempster's rule of combination. Therefore we take advantage of the fact famous from [11] saying that the commonality function $(Q_1 \oplus_D Q_2)$ corresponding to the basic assignment $(m_1 \oplus_D m_2)$ can easily be got as the pointwise product of commonality functions Q_1 and Q_2 corresponding to basic assignments m_1 and m_2, respectively. More precisely

$$(Q_1 \oplus_D Q_2)(A) = \Gamma^{-1} Q_1(A^{\downarrow K}) \cdot Q_2(A^{\downarrow L}),$$

where Γ is again a normalization constant, which is now computed

$$\Gamma = \sum_{A \subseteq \mathbb{X}_{K \cup L}} (-1)^{|A|+1} Q_1(A^{\downarrow K}) \cdot Q_2(A^{\downarrow L}).$$

From the definition of the combination operator for commonality functions, one can immediately see that the inverse removal operator must be defined for all $A \subseteq \mathbb{X}_{K \cup L}$

$$(Q_1 \ominus_D Q_2)(A) = \begin{cases} \Gamma^{-1} \frac{Q_1(A^{\downarrow K})}{Q_2(A^{\downarrow L})} & \text{if } Q_2(A^{\downarrow L}) > 0, \\ 0 & \text{otherwise,} \end{cases}$$

with

$$\Gamma = \sum_{A \subseteq \mathbb{X}_{K \cup L} : Q_2(A^{\downarrow L}) > 0} (-1)^{|A|+1} \frac{Q_1(A^{\downarrow K})}{Q_2(A^{\downarrow L})}.$$

So, following the results from [7], within D-S theory the proper operator of composition is defined

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

Its main disadvantage is its great computational complexity following, among others, from the fact that we do not know other way how to compute the composition $\triangleright_D$ of basic assignments than first transforming basic assignments $m_1, m_2, m_2^{\downarrow K \cap L}$ into the corresponding commonality functions, computing $Q_1 \triangleright_D$

$Q_2 = Q_1 \oplus_D Q_2 \ominus_D Q_2^{\downarrow K\cap L}$, and afterwards transforming the resulting composed commonality function back into the corresponding basic assignment.

One of the results from [7] says that the operator of composition $\triangleright_D$ is different from the one defined in [8], which we are going to introduce now. In what follows, the operator from [8] will be denoted $\triangleright_F$.

Consider two arbitrary basic assignments m_1 on $\mathbb{X}_K$ and m_2 on $\mathbb{X}_L$ ($K \neq \emptyset \neq L$) a *composition* $m_1 \triangleright_F m_2$ is defined for each $C \subseteq \mathbb{X}_{K\cup L}$ by one of the following expressions:

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

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

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

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

[c] in all other cases $(m_1 \triangleright_F m_2)(C) = 0$.

Let us note that similarly to $\triangleright_D$, also the operator $\triangleright_F$ can be expressed in the form of formula (1) but, naturally, with a different operator of combination. We will not need it in this paper, nevertheless let us mention for the interested reader that the corresponding operator $\oplus_F$ for m_1, m_2 on $\mathbb{X}_K, \mathbb{X}_L$, respectively, is defined by the following formula (for each $A \in \mathbb{X}_{K\cup L}$)

$$(m_1 \oplus_F m_2)(A) = \begin{cases} \Gamma^{-1} m_1(A^{\downarrow K}) m_2(A^{\downarrow L}) & \text{if } A = A^{\downarrow K} \bowtie A^{\downarrow L}, \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\Gamma = \sum_{A \subseteq \mathbb{X}_{K\cup L}: A = A^{\downarrow K} \bowtie A^{\downarrow L}} m_1(A^{\downarrow K}) \cdot m_2(A^{\downarrow L}).$$

Returning back to the main topic of this paper, let us summarize that in this section we have introduced two operators of composition $\triangleright_D$ and $\triangleright_F$. Though they differ from each other, as expressed in the following Proposition (for proofs see [8, 7]), both of them meet the basic properties required from an operator of composition.

Proposition 1 *Let m_1 and m_2 be basic assignments defined on $\mathbb{X}_K, \mathbb{X}_L$, respectively. Then both operators of composition $\triangleright_D$ and $\triangleright_F$ meet the following properties:*

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

The reader probably noticed that Property 2 guarantees that if $L \subseteq K$ then $m_1 \triangleright_D m_2 = m_1 \triangleright_F m_2$. It is really an easy task to show that the same equality holds true also when $K \cap L = \emptyset$. Nevertheless, not too much is known about other situations. It is clear that the above conditions are not necessary. Namely, the same equality holds true when one composes Bayesian basic assignments (i.e. basic assignments whose all focal elements are singletons). This is why we can formulate the first open problem.

Open Problem 1 *Is it possible to specify necessary and sufficient conditions under which $m_1 \triangleright_D m_2 = m_1 \triangleright_F m_2$?*

2 IPFP

In this section we will describe the Iterative Proportional Fitting Procedure with the help of the operator of composition. It can be applied to a system of basic assignments using any of the two operators of composition introduced in the previous section. This is why we use just the symbol $\triangleright$. It is important to realize, that for this computational process we need an operator possessing all the properties from Proposition 1, and we do not know any other operator meeting these properties.

Assume there is a system of n low-dimensional basic assignments $m_1, m_2, \ldots, m_n$ defined on $\mathbb{X}_{K_1}, \mathbb{X}_{K_2}, \ldots, \mathbb{X}_{K_n}$, respectively. During the computational process, an infinite sequence of basic assignments $\mu_0, \mu_1, \mu_2, \mu_3, \ldots$ is computed, each of them defined on $\mathbb{X}_{K_1\cup\ldots\cup K_n}$. In case this sequence is convergent, its limit is the result of this process.

Algorithm IPFP Define the starting basic assignment μ_0 on $\mathbb{X}_{K_1\cup K_2\cup\ldots\cup K_n}$.
Then compute

$$\begin{aligned}
\mu_1 &= m_1 \triangleright \mu_0 \\
\mu_2 &= m_2 \triangleright \mu_1 \\
\mu_3 &= m_3 \triangleright \mu_2 \\
&\vdots \\
\mu_n &= m_n \triangleright \mu_{n-1} \\
\mu_{n+1} &= m_1 \triangleright \mu_n \\
&\vdots
\end{aligned}$$

focal elements	m
$\{a_1\bar{a}_2a_3, \bar{a}_1a_2a_3\}$	0.2
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2a_3\}$	0.3
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2\bar{a}_3, \bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2a_3\}$	0.5

Table 1: Three-dimensional assignment m

$$\mu_{2n} = m_n \triangleright \mu_{2n-1}$$
$$\mu_{2n+1} = m_1 \triangleright \mu_{2n}$$
$$\vdots$$

As said in Introduction, when this algorithm is applied to probability measures, it has some nice and useful properties, most of which were proved by Csiszár in his famous paper [2]. So it is not surprising that the general properties formulated and proved here for belief functions (including the presented proofs) are based on the Csiszár's results.

Theorem 1 *If the sequence $\mu_0, \mu_1, \mu_2, \mu_3, \ldots$ computed by the Algorithm IPFP converges then the basic assignment*

$$\mu^* = \lim_{i\to+\infty} \mu_i$$

is a common extension of all $m_1, m_2, \ldots, m_n$, i.e.,

$$(\mu^*)^{\downarrow K_j} = m_j$$

for all $j = 1, \ldots, n$.

Proof. Consider any $j \in \{1, 2, \ldots, n\}$. From Property 2. of Proposition we get that m_j is marginal of all the assignments $\mu_j, \mu_{n+j}, \mu_{2n+j}, \mu_{3n+j}, \ldots$, and therefore m_j is marginal also to the limit of this subsequence

$$(\lim_{k\to+\infty} \mu_{kn+j})^{\downarrow K_j} = m_j.$$

From the basic course on mathematical analysis we know that if a sequence converges, then all their subsequences converge, too, and the limits are the same. Therefore, $(\mu^*)^{\downarrow K_j} = m_j$. □

2.1 IPFP with $\triangleright_F$

Example 1 Let us first illustrate and comment the process on a simple example. Consider a three-dimensional space $\mathbb{X}_{\{1,2,3\}}$, with $\mathbb{X}_i = \{a_i, \bar{a}_i\}$. To be sure that the considered system of two-dimensional basic assignments is consistent, i.e., that there exists their common extension, consider the three-dimensional assignment on $\mathbb{X}_{\{1,2,3\}}$ with three focal elements from Table 1. Its two-dimensional marginal assignments $m_1 = m^{\downarrow\{1,2\}}, m_2 = m^{\downarrow\{2,3\}}$ and $m_3 = m^{\downarrow\{1,3\}}$ are in Table 2.

	focal elements	values
m_1	$\{a_1\bar{a}_2, \bar{a}_1a_2\}$	0.2
	$\{a_1a_2, a_1\bar{a}_2\}$	0.3
	$\{a_1a_2, a_1\bar{a}_2, \bar{a}_1a_2, \bar{a}_1\bar{a}_2\}$	0.5
m_2	$\{a_2a_3\}$	0.2
	$\{a_2\bar{a}_3, \bar{a}_2a_3\}$	0.3
	$\{a_2a_3, a_2\bar{a}_3, \bar{a}_2a_3, \bar{a}_2\bar{a}_3\}$	0.5
m_3	$\{a_1a_3, \bar{a}_1a_3\}$	0.2
	$\{a_1a_3, a_1\bar{a}_3\}$	0.3
	$\{a_1\bar{a}_3, \bar{a}_1, a_3, \bar{a}_2\bar{a}_3\}$	0.5

Table 2: Consistent assignments m_1, m_2, m_3

The computational process starting with $\mu_0(A) = 1/255$ for all nonempty $A \subseteq \mathbb{X}_{\{1,2,3\}}$ is depicted in Table 3. We do not present here assignments μ_1 and μ_2, because they have 99, and 15 focal elements, respectively. Starting with μ_3 all the remaining computations concern only six focal elements represented by six rows of Table 3. Looking at this table the reader perhaps believes that the process converges, and that the limit assignment has eventually only four focal elements.

The convergence of the procedure in the previous example is not surprising because for $\triangleright_F$ we can use the ideas from the Csiszár's proof [2] to get the following theorem.

Theorem 2 *Consider a system of proper basic assignments $m_1, m_2, \ldots, m_n$ defined on $\mathbb{X}_{K_1}, \mathbb{X}_{K_2}, \ldots, \mathbb{X}_{K_n}$ and a proper basic assignment μ_0 on $\mathbb{X}_{K_1\cup\ldots\cup K_n}$. If there exists a proper basic assignment ν on $\mathbb{X}_{K_1\cup\ldots\cup K_n}$ such that ν is dominated by μ_0, and ν is a common extension of all $m_1, m_2, \ldots, m_n$, then the sequence $\mu_0, \mu_1, \mu_2, \mu_3, \ldots$ computed by the Algorithm IPFP with $\triangleright_F$ converges.*

The proof is based on the following auxiliary assertion.

Lemma 1 *Consider two basic proper assignments μ, ν on $\mathbb{X}_L$, and let $K \subseteq L$. Denote*

$$D(\nu\|\mu) = \sum_{A\subseteq\mathbb{X}_L:\mu(A)>0} \mu(A) \log \frac{\mu(A)}{\nu(A)}.$$

If ν dominates μ (i.e., $\nu(A) = 0 \Rightarrow \mu(A) = 0$) then

$$D(\nu\|\mu) = D(\nu\|\mu^{\downarrow K} \triangleright_F \nu) + D(\mu^{\downarrow K} \triangleright_F \nu\|\mu).$$

focal elements	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	μ_{100}	μ_{1000}
$\{a_1\bar{a}_2a_3, \bar{a}_1a_2a_3\}$	0.156	0.200	0.166	0.166	0.200	0.172	0.195	0.199
$\{a_1a_2a_3, a_1\bar{a}_2a_3, \bar{a}_1a_2a_3, \bar{a}_1\bar{a}_2a_3\}$	0.043	0.040	0.033	0.033	0.031	0.027	0.004	$4 \cdot 10^{-4}$
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2a_3\}$	0.146	0.146	0.300	0.211	0.211	0.300	0.293	0.299
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2a_3, a_1\bar{a}_2\bar{a}_3\}$	0.153	0.153	0.124	0.088	0.085	0.079	0.006	$7 \cdot 10^{-4}$
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2a_3, \bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2a_3\}$	0.250	0.230	0.187	0.250	0.234	0.210	0.250	0.250
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2\bar{a}_3, \bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2a_3, \bar{a}_1\bar{a}_2\bar{a}_3\}$	0.250	0.230	0.187	0.250	0.234	0.210	0.250	0.250

Table 3: $m_1 \triangleright_F \mu_0$

Proof.

$$
\begin{aligned}
D(\nu\|\mu) &= \sum_{A\subseteq \mathbb{X}_L:\mu(A)>0} \mu(A)\log\left(\frac{\mu(A)}{\nu(A)}\cdot\frac{(\mu^{\downarrow K}\triangleright_F \nu)(A)}{(\mu^{\downarrow K}\triangleright_F \nu)(A)}\right)\\
&= \sum_{A\subseteq \mathbb{X}_L:\mu(A)>0} \mu(A)\log\frac{\mu(A)}{(\mu^{\downarrow K}\triangleright_F \nu)(A)}\\
&\quad + \sum_{A\subseteq \mathbb{X}_L:\mu(A)>0} \mu(A)\log\frac{(\mu^{\downarrow K}\triangleright_F \nu)(A)}{\nu(A)}\\
&= D(\mu^{\downarrow K}\triangleright_F \nu\|\mu)\\
&\quad + \sum_{A\subseteq \mathbb{X}_L:\mu(A)>0} \mu(A)\log\frac{(\mu^{\downarrow K}\triangleright_F \nu)(A)}{(\nu^{\downarrow K}\triangleright_F \nu)(A)}.
\end{aligned}
$$

The last modification is based on Property 4 of Proposition.

Realize, now, that the last summation is performed over those $A \subseteq \mathbb{X}_L$ for which $\mu(A) > 0$, and therefore, due to the assumed dominance, $\nu(A) > 0$, too. Therefore, both $(\mu^{\downarrow K}\triangleright_F \nu)(A)$ and $(\nu^{\downarrow K}\triangleright_F \nu)(A)$ are computed according to case [**a**] of the respective definition getting

$$
\frac{(\mu^{\downarrow K}\triangleright_F \nu)(A)}{(\nu^{\downarrow K}\triangleright_F \nu)(A)} = \frac{\frac{\mu^{\downarrow K}(A^{\downarrow K})\cdot\nu(A)}{\nu^{\downarrow K}(A^{\downarrow K})}}{\frac{\nu^{\downarrow K}(A^{\downarrow K})\cdot\nu(A)}{\nu^{\downarrow K}(A^{\downarrow K})}} = \frac{\mu^{\downarrow K}(A^{\downarrow K})}{\nu^{\downarrow K}(A^{\downarrow K})}.
$$

So, we can proceed further in computation of $D(\nu\|\mu)$:

$$
\begin{aligned}
D(\nu\|\mu) &= D(\mu^{\downarrow K}\triangleright_F \nu\|\mu)\\
&\quad + \sum_{A\subseteq \mathbb{X}_L:\nu(A)>0} \mu(A)\log\frac{\mu^{\downarrow K}(A^{\downarrow K})}{\nu^{\downarrow K}(A^{\downarrow K})}\\
&= D(\mu^{\downarrow K}\triangleright_F \nu\|\mu)\\
&\quad + \sum_{\substack{B\subseteq \mathbb{X}_K\\ \nu(B)>0}} \sum_{\substack{A\subseteq \mathbb{X}_L:\nu(A)>0\\ A^{\downarrow K}=B}} \mu(A)\log\frac{\mu^{\downarrow K}(A^{\downarrow K})}{\nu^{\downarrow K}(A^{\downarrow K})}\\
&= D(\mu^{\downarrow K}\triangleright_F \nu\|\mu)\\
&\quad + \sum_{\substack{B\subseteq \mathbb{X}_K\\ \nu(B)>0}} \log\frac{\mu^{\downarrow K}(B)}{\nu^{\downarrow K}(B)} \sum_{\substack{A\subseteq \mathbb{X}_L:\nu(A)>0\\ A^{\downarrow K}=B}} \mu(A)\\
&= D(\mu^{\downarrow K}\triangleright_F \nu\|\mu)\\
&\quad + \sum_{B\subseteq \mathbb{X}_K:\nu(B)>0} \mu(B)\log\frac{\mu^{\downarrow K}(B)}{\nu^{\downarrow K}(B)},
\end{aligned}
$$

where the last modification is based on the formula for marginalization.

Regarding the fact that using analogous computations

$$
\begin{aligned}
D(\nu\|\mu^{\downarrow K}\triangleright_F \nu) &= \sum_{\substack{A\subseteq \mathbb{X}_L\\ (\mu^{\downarrow K}\triangleright_F\nu)(A)>0}} (\mu^{\downarrow K}\triangleright_F \nu)(A)\log\frac{(\mu^{\downarrow K}\triangleright_F \nu)(A)}{(\nu^{\downarrow K}\triangleright_F \nu)(A)}\\
&= \sum_{\substack{A\subseteq \mathbb{X}_L\\ (\mu^{\downarrow K}\triangleright_F\nu)(A)>0}} (\mu^{\downarrow K}\triangleright_F \nu)(A)\log\frac{\mu^{\downarrow K}(A^{\downarrow K})}{\nu^{\downarrow K}(A^{\downarrow K})}\\
&= \sum_{B\subseteq \mathbb{X}_K:\nu(B)>0} \mu(B)\log\frac{\mu^{\downarrow K}(B)}{\nu^{\downarrow K}(B)},
\end{aligned}
$$

we have finished the proof. □

Proof of Theorem 2. First notice that the function $D(\nu\|\mu)$ introduced in the previous Lemma is in fact the famous Kullback-Leibler divergence between two probability measures (let us stress that we assume that all the involved basic assignments are proper, because $\triangleright_F$ composition of two proper assignments is obviously also proper) defined on $2^{\mathbb{X}_L}$, which is known to be nonnegative, equals 0 if and only if $\nu = \mu$, and is finite if ν dominates μ. Moreover, since ν is assumed to be a common extension of all $m_1, m_2, \ldots, m_n$, it means that $\nu^{\downarrow K_j} = m_j$ for all $j = 1, 2, \ldots, n$.

So, following the idea of Csiszár, we can apply

Lemma 1 getting

$$D(\mu_0\|\nu) = D(\mu_0\|m_1 \triangleright_F \mu_0) + D(m_1 \triangleright_F \mu_0\|\nu),$$

where $m_1 \triangleright_F \mu_0 = \mu_1$ computed by Algorithm IPFP. Analogously,

$$D(\mu_1\|\nu) = D(\mu_1\|\mu_2) + D(\mu_2\|\nu),$$
$$D(\mu_2\|\nu) = D(\mu_2\|\mu_3) + D(\mu_3\|\nu),$$
$$\vdots$$

and therefore

$$D(\mu_0\|\nu) \geq \sum_{j=1}^{\infty} D(\mu_{j-1}\|\mu_j).$$

Since we assume that μ_0 dominates ν, $D(\mu_0\|\nu)$ is finite, and therefore

$$\lim_{j\to\infty} D(\mu_{j-1}\|\mu_j) = 0.$$

The required convergence of $\mu_0, \mu_1, \mu_2, \mu_3, \ldots$ follows directly from the fact that the last equality guarantees also that (for more details see [2])

$$\lim_{j\to\infty} \sum_{A\subseteq \mathbb{X}_{K_1\cup\ldots\cup K_n}} |\mu_{j-1}(A) - \mu_j(A)| = 0.$$ □

Example 2 Let us conclude this section with an example illustrating behavior of the Algorithm IPFP in case of an inconsistent system of basic assignments. It is clear that IPFP does not converge in this case, because, due to Theorem 1, otherwise it would have converged to a joint extension of the given assignments, which does not exist. However, based on our experiments, there exist converging subsequences. This phenomenon is known also from the probabilistic IPFP [14].

Let us consider three basic assignments m_1, m_2, and m_3 defined on $\mathbb{X}_{\{1,2\}}$, $\mathbb{X}_{\{2,3\}}$, $\mathbb{X}_{\{1,3\}}$, respectively, where, again, $\mathbb{X}_i = \{a_i, \bar{a}_i\}$. The focal elements of these assignments as well as the respective values are in Table 4.

	focal elements	values
m_1	$\{\bar{a}_1a_2\}$	0.55
	$\{a_1\bar{a}_2, \bar{a}_1a_2\}$	0.40
	$\{a_1a_2, \bar{a}_1a_2, \bar{a}_1\bar{a}_2\}$	0.05
m_2	$\{a_2a_3\}$	0.63
	$\{a_2a_3, a_2\bar{a}_3, \bar{a}_2a_3\}$	0.22
	$\{a_2a_3, a_2\bar{a}_3, \bar{a}_2\bar{a}_3\}$	0.15
m_3	$\{\bar{a}_1a_3\}$	0.65
	$\{a_1a_3, \bar{a}_1a_3, \bar{a}_1\bar{a}_3\}$	0.35

Table 4: Inconsistent assignments m_1, m_2, m_3

Now, let us perform the IPFP process with μ_0 that is the same as in Example 1: $\mu_0(A) = 1/255$ for all nonempty $A \subseteq \mathbb{X}_{\{1,2,3\}}$. A part of the computational process is depicted in Table 5.

In this situation, the beginning of the process is not interesting. But after a several cycles, we can see that the iteration process goes through cyclical changes. From this example we can see that there are three convergent subsequences, namely

$$\mu_1, \mu_4, \mu_7, \ldots, \mu_{3k+1}, \ldots$$
$$\mu_2, \mu_5, \mu_8, \ldots, \mu_{3k+2}, \ldots$$
$$\mu_3, \mu_6, \mu_9, \ldots, \mu_{3k}, \ldots$$

In all our computational experiments it appeared that the length of the cycle which the process goes through corresponds to the number of basic assignments entering the computational process, and that the respective subsequences converged.

2.2 IPFP with $\triangleright_D$

Let us say at the very beginning of this section that considering the operator $\triangleright_D$ leads to many open problems. One of the reasons is connected with the computational complexity of this operator. Namely, computational complexity of composition operators is, naturally, closely connected with the number of focal elements to be enumerated. As a rule, D-operator produces a higher number of focal elements in comparison with F-operator. Moreover, in case of F-operator the enumeration of a value of a basic assignment for each focal element is got as a product of the respective projections of the focal element (i.e. a product of only two numbers), for D-operator one needs to process all the supersets of the respective projections. Thus, we can apply the IPFP Algorithm with D-operator only to very simple examples and even for them we cannot compute too long sequences $\mu_0, \mu_1, \mu_2, \mu_3, \ldots$. Other difficulties connected with application of this operator of composition will be formulated as open problems. The first one is connected with the fact, that in contrast to $\triangleright_F$, composition $\triangleright_D$ of two proper basic assignments need not be proper - it can achieve negative values.

Open Problem 2 *What are the necessary and sufficient conditions guaranteeing that $\triangleright_D$ composition of two proper assignments is also proper?*

Example 3 Consider first the same system of three consistent basic assignments as in Example 1, and start the computational process again with $\mu_0(A) = 1/255$ for all nonempty $A \subseteq \mathbb{X}_{\{1,2,3\}}$. Assignments μ_1 and μ_2 have now 99, and 70 focal elements, respectively. Starting with μ_3 all the remaining computations concern 44 focal elements, and nearly half

focal elements	μ_{13}	μ_{14}	μ_{15}	μ_{16}	μ_{17}	μ_{18}	μ_{43}	μ_{44}	μ_{45}
$\left\{\begin{matrix} a_1a_2a_3, \bar{a}_1a_2a_3, \\ \bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2\bar{a}_3 \end{matrix}\right\}$	0.049	0.150	0.142	0.049	0.150	0.142	0.050	0.150	0.142
$\left\{\begin{matrix} a_1a_2a_3, \bar{a}_1a_2\bar{a}_3, \\ \bar{a}_1\bar{a}_2a_3 \end{matrix}\right\}$	0.001	$2 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$7 \cdot 10^{-5}$	$5 \cdot 10^{-5}$	$5 \cdot 10^{-5}$	10^{-10}	10^{-10}	10^{-10}
$\left\{\begin{matrix} a_1a_2a_3, \bar{a}_1a_2a_3, \\ \bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2a_3 \end{matrix}\right\}$	0.001	$2 \cdot 10^{-4}$	$2 \cdot 10^{-4}$	$7 \cdot 10^{-5}$	$5 \cdot 10^{-5}$	$5 \cdot 10^{-5}$	10^{-10}	10^{-10}	10^{-10}
$\left\{\begin{matrix} a_1\bar{a}_2a_3, \bar{a}_1a_2a_3, \\ \bar{a}_1a_2\bar{a}_3 \end{matrix}\right\}$	0.400	0.219	0.208	0.400	0.219	0.208	0.400	0.220	0.208
$\{\bar{a}_1a_2a_3\}$	0.550	0.630	0.650	0.550	0.630	0.650	0.550	0.630	0.650

Table 5: IPFP $\triangleright_F$: inconsistent marginals

focal elements	μ_3	μ_4	μ_5	μ_6	μ_7	μ_{100}	μ_{1000}
$\{a_1\bar{a}_2a_3, \bar{a}_1a_2a_3\}$	0.020	0.030	0.033	0.031	0.039	0.085	0.095
$\{a_1a_2a_3, a_1\bar{a}_2a_3, \bar{a}_1a_2a_3, \bar{a}_1\bar{a}_2a_3\}$	0.017	0.042	0.046	0.042	0.047	0.031	0.010
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2a_3\}$	0.141	0.208	0.233	0.168	0.203	0.294	0.299
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2a_3, a_1\bar{a}_2\bar{a}_3\}$	0.103	0.152	0.140	0.101	0.122	0.014	10^{-4}
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2\bar{a}_3, \bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2a_3\}$	0.097	0.226	0.208	0.232	0.260	0.413	0.476
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2\bar{a}_3, \bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2a_3, \bar{a}_1\bar{a}_2\bar{a}_3\}$	0.097	0.226	0.208	0.232	0.260	0.413	0.476
$\{a_1a_2\bar{a}_3, a_1\bar{a}_2\bar{a}_3, \bar{a}_1a_2\bar{a}_3\}$	−0.047	−0.034	−0.045	−0.090	−0.051	−0.190	−0.228
$\{\bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2\bar{a}_3\}$	−0.021	−0.020	−0.015	−0.001	0.004	0.012	0.001

Table 6: IPFP $\triangleright_D$: converging sequence for consistent marginals

of them have negative values. After a thousand of iterative steps the changes are so small that we can take μ_{1000} as a limit of the computational process. In agreement with Theorem 1 we can see that $m_1 \doteq m_{1000}^{\downarrow\{1,2\}}, m_2 \doteq m_{1000}^{\downarrow\{2,3\}}$ and $m_3 \doteq m_{1000}^{\downarrow\{1,3\}}$.

A part of the computational process is depicted in Table 6. We selected 8 focal elements, the first 6 of them correspond to those from Table 3, the other 2 are chosen to present examples of focal elements with negative values. Observe that the last focal element switched its value from a negative one to a positive one during the IPFP.

Example 4 It shows up that in contrast to the application of $\triangleright_F$, the Algorithm IPFP with $\triangleright_D$ need not converge for a consistent system of marginal basic assignments. As an example consider the 3-dimensional assignment m from Table 7 and its marginals $m_1 = m^{\downarrow\{1,2\}}, m_2 = m^{\downarrow\{2,3\}}, m_3 = m^{\downarrow\{1,3\}}$. With μ_0 as in the previous examples, the sequence $\mu_0, \mu_1, \mu_2, \mu_3, \ldots$ computed by the Algorithm IPFP does not converge - it stabilizes in a loop of length 6 after approximately 560 iterations (i.e. $\mu_{601} = \mu_{607}, \mu_{602} = \mu_{608}, \ldots$). The strange behavior of this process is visible from Table 8, where a selected part of focal elements are presented. There are two phenomena that are in a way surprising. First, it is the length of the cycle (6), and the fact that even focal elements may variate during the cycle.

focal elements	m
$\{\bar{a}_1\bar{a}_2\bar{a}_3\}$	0.225
$\{a_1\bar{a}_2\bar{a}_3, \bar{a}_1a_2a_3\}$	0.126
$\{\bar{a}_1a_2\bar{a}_3, \bar{a}_1\bar{a}_2\bar{a}_3\}$	0.594
$\{\bar{a}_1\bar{a}_2a_3, \bar{a}_1\bar{a}_2\bar{a}_3\}$	0.024
$\{a_1\bar{a}_2a_3, a_1\bar{a}_2\bar{a}_3, \bar{a}_1\bar{a}_2a_3, \bar{a}_1\bar{a}_2\bar{a}_3\}$	0.031

Table 7: Three-dimensional assignment m

Open Problem 3 *Under what conditions does the sequence $\mu_0, \mu_1, \mu_2, \mu_3, \ldots$ computed by the Algorithm IPFP with $\triangleright_D$ converge? When is the limit assignment proper?*

3 Summary and Conclusions

Using two different operators of composition for belief functions that were studied in [6, 7], we designed two versions of the iterative procedure presented as Algorithm IPFP. If they converge, both of these algorithms yield basic assignments that have the input low-dimensional assignments for their marginals. But this is perhaps the only property common to both of them. Even in case that both the algorithms converge, the results may be different. In fact, we conjecture that these algorithms yield the same results

focal elements	μ_{601}	μ_{602}	μ_{603}	μ_{604}	μ_{605}	μ_{606}	μ_{607}	μ_{608}	μ_{609}
$\{\bar{a}_1 a_2 a_3\}$	-0.008	0	-0.047	0.036	0	-0.052	-0.008	0	-0.047
$\{a_1 \bar{a}_2 a_3\}$	-0.023	0.008	0.047	0.008	-0.007	0.052	-0.023	0.008	0.047
$\{a_1 \bar{a}_2 \bar{a}_3\}$	0.008	-0.003	-0.123	-0.036	0.006	0.033	0.008	-0.003	-0.123
$\{\bar{a}_1 \bar{a}_2 a_3\}$	-0.076	0.012	0.348	-0.038	-0.051	0.279	-0.076	0.012	0.348
$\{\bar{a}_1 \bar{a}_2 \bar{a}_3\}$	0.227	0.205	0	0.232	0.242	0	0.227	0.205	0
$\{a_1 \bar{a}_2 \bar{a}_3, \bar{a}_1 a_2 a_3\}$	0.110	0.082	0.157	0.066	0.071	0.157	0.110	0.082	0.157
$\{\bar{a}_1 a_2 a_3, \bar{a}_1 \bar{a}_2 \bar{a}_3\}$	0	0.043	0	0	0.054	0	0	0.043	0
$\{a_1 \bar{a}_2 a_3, \bar{a}_1 \bar{a}_2 a_3\}$	0.058	-0.004	0.055	0.050	0.001	0.055	0.058	-0.004	0.055
$\{a_1 \bar{a}_2 a_3, a_1 \bar{a}_2 \bar{a}_3\}$	0.076	0.034	0	0.038	0.046	0	0.076	0.034	0
$\{a_1 \bar{a}_2 \bar{a}_3, \bar{a}_1 \bar{a}_2 a_3\}$	0.044	0.020	-0.031	0.007	0.008	-0.031	0.044	0.020	-0.031
$\{\bar{a}_1 \bar{a}_2 a_3, \bar{a}_1 \bar{a}_2 \bar{a}_3\}$	0.044	0.020	-0.031	0.007	0.008	-0.031	0.044	0.020	-0.031
$\{a_1 \bar{a}_2 \bar{a}_3, \bar{a}_1 a_2 \bar{a}_3\}$	0.015	0.016	0.140	0.059	0.058	0.022	0.015	0.016	0.140
$\{\bar{a}_1 a_2 \bar{a}_3, \bar{a}_1 \bar{a}_2 \bar{a}_3\}$	0.535	0.576	0.593	0.542	0.535	0.505	0.535	0.576	0.593
$\{a_1 \bar{a}_2 \bar{a}_3, \bar{a}_1 \bar{a}_2 \bar{a}_3\}$	0.031	0.007	-0.140	0.031	0.033	-0.022	0.031	0.007	-0.140
$\left\{\begin{matrix} a_1 \bar{a}_2 a_3, a_1 \bar{a}_2 \bar{a}_3, \\ \bar{a}_1 \bar{a}_2 a_3, \bar{a}_1 \bar{a}_2 \bar{a}_3 \end{matrix}\right\}$	-0.044	-0.020	0.031	-0.007	-0.008	0.031	-0.044	-0.020	0.031

Table 8: IPFP $\triangleright_D$: non-converging sequence for consistent marginals

only in degenerate situations. As a rule, application of $\triangleright_D$ yields basic assignments with greater number of focal elements (compare Examples 1 and 3).

The algorithm employing $\triangleright_F$ manifests some of the nice properties of the probabilistic IPFP: its convergence is guaranteed for consistent systems of low-dimensional assignments. Moreover, its significantly lower computational complexity predestinates this version of the algorithm to practical applications. Another its advantage follows from the fact that if the input assignments are proper then the resulting basic assignment is also proper, which is not true for the Algorithm based on $\triangleright_D$. For example, when we randomly generated three-dimensional basic assignments, and applied the Algorithm IPFP with $\triangleright_D$ to their two-dimensional marginals, only about every fifteens solution was proper.

As it was highlighted in one of the referee reports, the application of the IPFP procedure may be extended beyond probability theory to other topics as, for example, that described in [9]. In fact, as the title of the paper suggests, the authors see several ways how to prolong the research in the field.

Acknowledgements

This work was supported by GAČR: the first author under the grant no. 403/12/2175, the second one under the grant no. 13-20012S.

References

[1] P. Cheeseman. In defence of probability. In *Proc. 8th Int. Joint Conf. on AI (IJCAI'85).* Los Angeles, CA, pp 1002–1009, 1985.

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

[3] 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.

[4] R. Jiroušek. A note on local computations in Dempster-Shafer theory of evidence. In *Proc. of the 7th Symp. on Imprecise Probabilities and Their Applications,* STUDIA Universitatsverlag, Innsbruck, pp. 219–227, 20011.

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

[6] R. Jiroušek and P. P. Shenoy. Compositional models in valuation-based systems. In *Belief Functions: Theory and Applications*, T. Denoeux and M.-H. Masson, eds., Advances in Intelligent and Soft Computing 164, pp 221–228, Heidelberg, Springer, 2012.

[7] R. Jiroušek and P. P. Shenoy. Compositional models in valuation-based systems. In print *Int. J. Approx. Reasoning*, doi: 10.1016/j.ijar.2013.02.002.

[8] 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.

[9] E. Miranda, M. Zaffalon. Coherence graphs. *Artificial Intelligence,* 173(1), 104-144, 2009.

[10] R. D. Shachter. Intelligent probabilistic inference. In: *Uncertainty in AI.* L. N. Kanal, J. F. Lemmer, eds. North Holland, Amsterdam, 1986.

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

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

[13] D. J. Spiegelhater. Probabilistic reasoning in predictive expert systems. In: *Uncertainty in Artificial Intelligence (UAI-85).* L. N. Kanal, J. F. Lemmer, eds. North Holland, Amsterdam, pp. 47–67, 1985.

[14] J. Vomlel. Integrating inconsistent data in a probabilistic model. *J. of Applied Non-Classical Logics,* 14, 3, pp. 367–386, 2004.