Approximate Evidential Reasoning Using Local Conditioning and Conditional Belief Functions

Van Nguyen
Defence Science and Technology Group
van.nguyen5@dst.defence.gov.au

Abstract

We propose a new message-passing belief propagation method that approximates belief updating on evidential networks with conditional belief functions. By means of local conditioning, the method is able to propagate beliefs on the original multiply-connected network structure using local computations, facilitating reasoning in a distributed and dynamic context. Further, by use of conditional belief functions in the form of partially defined plausibility and basic plausibility assignment functions, belief updating can be efficiently approximated using only partial information of the belief functions involved. Experiments show that the method produces results with high degree of accuracy whilst achieving a significant decrease in computational and space complexity (compared to exact methods).

1 INTRODUCTION

Network-based approaches for reasoning with uncertainty feature a graphical representation of the knowledge base (commonly a directed acyclic graph) equipped with bi-directional belief propagation, facilitating reasoning in a coherent manner. Such networks dynamically reflect the current state of knowledge through the process of *belief updating* where the impact of an observation presented at a node is propagated to other nodes throughout the network.

This work is concerned with belief updating in evidential networks with conditional belief functions such as ECNs [Xu and Smets1996] and DEVNs [Yaghlane and Mellouli2008] towards their practical application, motivated by a computer network defence (CND) system. Such networks are formulated based on the theory of belief functions (BF theory) [Shafer1976], specifically in the framework of Transferable Belief Models (TBM) [Smets and Kennes1994] which provides a greater expressive power to handle more general forms of knowledge, in particular incomplete knowledge, a vital requirement in our application. In particular, it is crucial to be able to distinguish between conflicting information, missing information and statistical variability (see [Dubois and Prade2009] for further discussion) in order to act appropriately; for instance, deploying additional sensors in the face of incomplete knowledge, or mitigating against adversary manipulation or faulty sensors in the case of conflicting information.

ECNs and DEVNs represent the relations between variables in the form of conditional belief functions which generally consume less space, and provide a more intuitive formalism to encode knowledge (albeit less generic) than joint belief functions on the product space as in well-known valuation networks [Shenoy1993]. Belief updating in networks with belief functions can be efficiently performed using local computations with message-passing mechanisms, provided that the network is singly-connected at *execution* time. This can be achieved by imposing a tree-structure (e.g., [Shenoy and Shafer1986, Shafer, Shenoy, and Mellouli1987]), or simplifying the structure of certain networks using partial dependency (e.g., [Xu and Smets1996]), or generation of a secondary structure for the network, such as a binary join tree (e.g., Figure 1b), using graph-theoretic techniques (e.g., [Shenoy1997, Yaghlane and Mellouli2008]). These methods assume knowledge of the network and/or involve significant pre-processing before belief updating can occur. In the context of our CND system, the reasoning networks of concern are a federation/integration of distributed subnetworks of knowledge dynamically emerging and submerging through activation/deactivation of different sets of sensors, actuators and tasks as demanded by the changing situation. As such not only is the topology and the full knowledge of the reasoning network not available prior to execution, but they may change at run-time.

In this paper, we propose a message-passing belief propagation method that does not assume/require the reasoning network to be singly-connected at and during execution. More specifically, the method operates directly on the *original* multiply-connected network (by means of local conditioning), and within this framework approximate computations using only partial information of the belief functions to be combined and propagated (by means of conditional belief functions represented in the form of partially defined plausibility and basic plausibility assignment functions). The method is thereby adaptive to changes in the network topology while associated with a significant reduction in space and computational complexity. We will present BF theory and networks using conditional belief functions in the next section. In Section 3, we briefly discuss local conditioning in Bayesian networks and how the technique can be extended to our networks of interest. In Section 4, we present our new approximate belief updating method that combines local conditioning with conditional belief functions. Discussions and experimental results are provided in Section 5 before concluding in Section 6.

2 BF THEORY AND NETWORKS OF CONDITIONAL BELIEF FUNCTIONS

Given a frame of discernment Θ, a finite nonempty set of all possible hypotheses, BF theory assigns belief mass to the elements of its powerset 2^{Θ}, effectively representing incompleteness by means of disjunctive sets. Formally, an (unnormalised) *mass function* over Θ is defined as a mapping $\mathrm{m} : 2^{\Theta} \to [0, 1]$ such that $\sum_{A \subseteq 2^{\Theta}} \mathrm{m}(A) = 1$. Those subsets A where $\mathrm{m}(A) > 0$ are called *focal elements*; with $\mathrm{m}(A) = 0$ whenever $|A| > 1$ representing a *Bayesian* belief function, and $\mathrm{m}(\Theta) = 1$ representing total ignorance. A mass function can be equivalently represented in the form of a *belief function* (bel), a *plausibility function* (pl) and a *commonality function* (q) with different semantics[1], given respectively as $\mathrm{bel}(A) = \sum_{B \subseteq A, B \neq \emptyset} \mathrm{m}(B)$, $\mathrm{pl}(A) = \mathrm{bel}(\Theta) - \mathrm{bel}(\bar{A})$ and $\mathrm{q}(A) = \sum_{A \subseteq B \subseteq \Omega} \mathrm{m}(B)$, where $\bar{A} = \Theta - A$. If there are multiple distinct pieces of evidence defined on Θ such as m_1 and m_2, their combined impact is computed through the conjunctive rule of combination (CRC) $\textcircled{\cap}$: $\forall A \subseteq \Omega, m_{1 \textcircled{\cap} 2}(A) = \sum_{B,C \subseteq \Omega, B \cap C = A} m_1(B) m_2(C)$, or equivalently $q_{1 \textcircled{\cap} 2}(A) = q_1(A) q_2(A)$ (2.1).

In his work formulating BF theory from a geometric perspective, [Cuzzolin2012] proposed the notion of *basic plausibility assignment* (or b.pl.a) as the Moebius inverse $\mu_{pl} : 2^{\Omega} \to \Re$ of a plausibility function pl such that:

$$\mu_{pl}(A) = \sum_{B \subseteq A} (-1)^{|A \setminus B|} pl(B) \quad (2.2)$$

$$pl(A) = \sum_{B \subseteq A} \mu_{pl}(B). \quad (2.3)$$

We now show that distinct pl functions can be combined by means of μ_{pl}, which plays a major role in this work.

Lemma 2.1 *: Let $pl_1, \ldots, pl_n$ be distinct pieces of evidence on Θ, their combined impact on Θ is computed by means of their corresponding $\mu_{pl_1}, \ldots, \mu_{pl_n}$ as: $\forall A \subseteq \Theta$,*

$$\mu_{pl_1 \textcircled{\cap} \ldots \textcircled{\cap} n}(A) = (-1)^{(|A|+1)(n-1)} \prod_{k=1}^{n} \mu_{pl_k}(A). \quad (2.4)$$

[1] m, bel, pl and q can all be referred to as *belief functions*.

Proof By iteratively substituting terms in the RHS of Eq. (2.2) with $pl(A) = \sum_{B \subseteq A} (-1)^{|B|+1} q(B)$ [Shafer1976] for each subset A with increasing size $|A|$, we can derive $q(A) = (-1)^{|A|+1} \mu(A)$ (2.5). Substituting q in Eq. (2.1) with μ according to Eq. (2.5) results in Eq. (2.4). ∎

When distinct m^{Θ} and m^{Ω} are expressed on different frames, e.g., Θ and Ω, [Smets1994] devised a formalism where the relation between Θ and Ω is represented by means of *conditional belief functions* (cbfs) $pl^{\Omega}[\theta_i](\omega)$ for each $\omega \subseteq \Omega$ given $\theta_i \in \Theta$. To allow for belief propagation in the presence of incomplete knowledge (e.g., the prior knowledge of Θ can be incomplete or even vacuous), [Smets1994] proposed the *Disjunctive Rule of Combination* (DRC), and conversely the *Generalised Bayes' Theorem* (GBT), to compute $pl^{\Omega}[\theta](\omega)$ and $pl^{\Theta}[\omega](\theta)$, respectively, for any $\theta \subseteq \Theta$, $\omega \subseteq \Omega$, from those $pl^{\Omega}[\theta_i](\omega)$:

$$pl^{\Omega}[\theta](\omega) = pl^{\Theta}[\omega](\theta) = 1 - \prod_{\theta_i \in \theta} (1 - pl^{\Omega}[\theta_i](\omega)). \quad (2.6)$$

As Eq. (2.6) is derived from the *principle of minimal commitment* (MCP) [Smets1994], the cbfs between any two nodes constructed in this way are said hereafter to be *MCP-compatible*. By Eq. (2.6), the impacts of m^{Θ} and m^{Ω} can be propagated to each other frame through:

$$pl^{\Omega}(\omega) = \sum_{\theta \subseteq \Theta} m^{\Theta}(\theta) pl^{\Omega}[\theta](\omega), \forall \omega \subseteq \Omega, \quad (2.7)$$

$$pl^{\Theta}(\theta) = \sum_{\omega \subseteq \Omega} m^{\Omega}(\omega) pl^{\Omega}[\omega](\theta), \forall \theta \subseteq \Theta. \quad (2.8)$$

The DRC and GBT allows a knowledge base to be constructed as a *directed acyclic graph* where nodes represent variables, and the directed edge between a node Ω and its parent Θ represents their relationship, initially described by the conditional belief functions $pl^{\Omega}[\theta_i](\omega), \forall \theta_i \in \Theta, \forall \omega \subseteq \Omega$. When a node receives observation in the form of a belief function, the impact will be propagated throughout the network to reflect the current state of knowledge. When the network or its secondary structure is singly-connected, belief propagation can be performed using a message-passing mechanism similar to that of Bayesian networks (see [Pearl1988]) with (i) beliefs propagated between nodes using DRC and GBT, and (ii) beliefs received at the same node combined using CRC (see [Xu and Smets1996, Yaghlane and Mellouli2008] for further discussion). When all the knowledge involved is complete, belief propagation using the DRC and GBT coincide with that of Bayesian reasoning (see [Smets1994]).

3 BELIEF UPDATING

Message-passing algorithms that compute beliefs directly on the original network have been devised for Bayesian networks (BNs), most notably the *loopy belief propagation* (LBP) and *(local) conditioning* algorithms (LC). LBP [Pearl1988] iteratively propagates beliefs as if the network is a poly-tree until convergence. Though providing a simple and often efficient mechanism for belief propagation, the results from LBP are theoretically known (e.g., in [Pearl1988]), and empirically verified (e.g., in [Weiss1997]) to not represent the posterior marginal of nodes which is of primary concern in belief updating. Thus LBP is not considered in this work. *(Global) conditioning* [Pearl1988] is based on the fact that instantiation of certain nodes (or a *cutset*) in the network will break the loop(s) containing them, rendering the network singly-connected. *Local conditioning* (LC) [Díez1996, Fay and Jaffray2000] improves on the original global method by ensuring that only nodes within a loop are conditioned on the variable that breaks the loop, and allows the impact from multiple observations to be simultaneously propagated. Application of LC in BNs was described in detail in [Díez1996, Fay and Jaffray2000]. In essence, LC is able to integrate belief propagation with the process of breaking loops and forming a tree associated with the network. When a loop is discovered, the method breaks the loop by splitting the conditioning variable into the original and its phantom instances and assigning the variable to every node within the loop (as illustrated in Figure 1c, the loop 3-4-5-6-7-3 is broken by 5, with 4 connected to 'phantom 5', and 3, 4, 6, 7 conditioned on 5). Once the cutset is found, belief propagation with LC in BNs is then similar to the original message-passing algorithm, except that each message now carries the unconditional probabilities on the product space of a node within the loop(s) and its conditioning variables, and no normalisation is performed. It is not hard to prove that LC can be directly extended and applied to networks with conditional belief functions by using the ballooning extension technique [Smets and Kennes1994] to transform the conditional belief functions between a node and its parent into the corresponding belief functions on their product space, and beliefs can be propagated in a classical way. However, this would diminish the original purpose of these types of network, and significantly increase the computational complexity and memory consumption (i.e., the space complexity required to represent a belief function on the product space with n variables of m states would be $\mathcal{O}(2^{m^n})$, rather than $\mathcal{O}(2^{mn})$ as in the case of conditional belief functions).

Our question therefore is *whether it is possible to apply local conditioning directly with conditional belief functions*. Such an approach would encounter two major obstacles. On the one hand, unlike a BN where conditional probabilities are just the normalised version of the unconditional ones, conditional belief functions correspond to only a certain portion of the underlying unconditional (joint) belief functions. On the other hand, while knowledge of the local cutset in a BN is always complete (a necessary condition for loop elimination), such knowledge in networks with cbfs can be incomplete (i.e., the belief function associated with the cutset can be non-Bayesian) and the MCP-compatibility required for DRC/GBT to propagate beliefs in this case is generally not met. The former demands combination of belief functions defined on arbitrary domains which are partially known, while the latter necessitates the devising of an extended version of DRC/GBT without the MCP compatibility assumption.

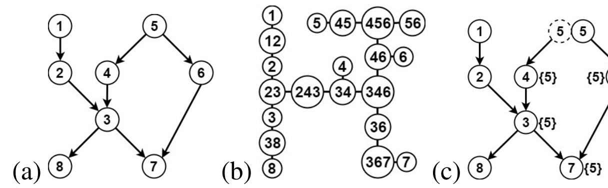


Figure 1: An example network (a), in the form of a binary join tree (adapted from [Yaghlane and Mellouli2008]) (b), or configured for local conditioning (c).

These obstacles will be addressed in Section 4. Specifically, we introduce the notion of (i) partially defined pl function, pl_{r}, that provides a compact and simultaneous encoding of multi-directional cbfs between a group of nodes of interest (e.g., a node and its parents/local cutset); and (ii) its corresponding partially defined $b.pl.a$ function, $\mu_{pl_{\mathrm{r}}}$, that serves as the computational construct facilitating efficient computations. We next devise efficient methods that allow direct combination and propagation of beliefs represented in the form of pl_{r} by means of $\mu_{pl_{\mathrm{r}}}$. The method $\circledcirc_r$ allows exact combination when such pl_{r} are associated with disjoint sets of conditioning variables (see Theorem 4.5). When those pl_r are associated with overlapping sets of conditioning variables, $\circledcirc_r$ enables exact computation of a particular portion, $\mathscr{C}_{pl_r}$, of their combined impact (see Theorem 4.6 Step 1). Based on $\mathscr{C}_{pl_r}$, the entire combined impact can be reconstructed when necessary; either by means of interval approximation as in Theorem 4.6 Step 2, or point estimation using a parameterised version of DRC/GBT. In order to discuss the proposed belief propagation method, we assume that the cutset has been determined and the network has been configured accordingly (e.g., see Figure 1c).

4 BELIEF UPDATING WITH LC & CBFS

It is sufficient to define pl_{r} and $\mu_{pl_{\mathrm{r}}}$ with respect to any three variables X, Y and Z. The following notations are adopted from [Yaghlane, Smets, and Mellouli2002b]:

- XYZ denotes the Cartesian product $X \times Y \times Z$;
- x_i, y_j, z_k denote elements, and x, y, z subsets of X, Y, Z;
- (x, y, z) denotes $\{(x_i, y_j, z_k) : x_i \in x, y_j \in y, z_k \in z\}$ for some x, y and z;
- pl^{Ω} denotes the plausibility function on Ω, $pl^{\Omega\downarrow X}$ the marginal of pl^{Ω} on X, and $pl^{\Omega}[y,z]^{\downarrow X}$ the plausibility function on X that results from conditioning pl^{Ω} on (y, z) and then marginalised on X[2];
- $x^{\uparrow XY}$ is the cylindrical extension of $x \subseteq X$ on XY : $x^{\uparrow XY} = (x, Y)$, and $\omega^{\downarrow X}$ is the projection of $\omega \subseteq \Omega$ on X: $\omega^{\downarrow X} = \{x_i : x_i \in X, x_i^{\uparrow \Omega} \cap \omega \neq \emptyset\}$.

[2]Note that conditioning and marginalisation are not commutative, e.g., $pl^{\Omega}[y,z]^{\downarrow X} \neq pl^{\Omega\downarrow X}[y,z]$.

4.1 PARTIALLY DEFINED pl_{r} AND $\mu_{pl_{\mathrm{r}}}$

Definition 1 *Let* r_Ω *be the set of subsets of* $\Omega = XYZ$ *consisting of* (x,y,z) *for all* $x \subseteq X$, $y \subseteq Y$, *and* $z \subseteq Z$. *A* partially defined pl_{r}^{Ω} *is a plausibility function on* Ω *whose value is known for only those subsets of* Ω *in* r_Ω.

Such r_Ω can be considered the set of *hyper-rectangle* subsets of Ω. Each pl_{r}^{Ω} is associated with a set of pl^{Ω} on Ω (denoted as $\mathscr{P}_{\mathrm{r}}^{\Omega}$) such that $pl^{\Omega}(\omega) = pl_{\mathrm{r}}^{\Omega}(\omega)$, $\forall \omega \in \mathrm{r}_\Omega$; and pl_{r} defined on a single variable corresponds to its marginal (e.g., $pl_{\mathrm{r}}^{X} \equiv pl^{X}$). By extending Lemma 1 and 2 in [Yaghlane, Smets, and Mellouli2002a] to a multivariate context, Lemma 4.1 is obtained for all $x \subseteq X$, $y \subseteq Y$, $z \subseteq Z$:

Lemma 4.1 *For any given* pl_{r}^{Ω} *defined on* Ω, *we have:*

$$pl_{\mathrm{r}}^{\Omega}(x,y,z) = pl_{\mathrm{r}}^{\Omega}[z]^{\downarrow XY}(x,y) = pl_{\mathrm{r}}^{\Omega}[y,z]^{\downarrow X}(x),$$
$$pl_{\mathrm{r}}^{\Omega\downarrow XY}(x,y) = pl_{\mathrm{r}}^{\Omega}(x,y,Z), pl_{\mathrm{r}}^{\Omega\downarrow X}(x) = pl_{\mathrm{r}}^{\Omega}(x,Y,Z).$$

As seen above, the collection of conditional belief functions allowing evidence to be propagated between X, Y and Z in a canonical manner corresponds to the partially defined pl_{r}^{Ω} on Ω. Thus, pl_{r}^{Ω} can be considered a compact encoding that simultaneously represents the conditional relationship between any nodes of interest, not only between child and parent nodes as presented in Section 2. In singly-connected networks, propagated beliefs are in their marginal form (Eqs. (2.7) and (2.8)). In multiply-connected networks under local conditioning, the beliefs sent and received by a node are potentially conditioned on some variables, represented here in the form of pl_{r}, necessitating methods to combine and propagate those pl_{r}. Combination of (completely defined) pl functions is possible by means of other functions as shown in Section 2. However, since Lemma 4.1 does not hold for any functions other than pl (e.g., $m[x](y) \neq m(x,y) \neq m[y](x)$), attempts to derive a method to combine beliefs in their partial form pl_{r} by means of such functions would result in a significant computational and spatial overhead. In order to facilitate efficient and simultaneous combination of pl_{r} functions, we propose for any given pl_{r} its dual representation $\mu_{pl_{\mathrm{r}}}$.

Definition 2 *For any given* pl_{r}^{Ω} *on* Ω, *the corresponding* partially defined $b.pl.a$ function $\mu_{pl_{\mathrm{r}}^{\Omega}}$ *is defined by:*

$$\mu_{pl_{\mathrm{r}}^{\Omega}}(x,y,z) = \textstyle\sum_{\omega'^{\downarrow X}=x, \omega'^{\downarrow Y}=y, \omega'^{\downarrow Z}=z} \mu_{pl^{\Omega}}(\omega') \quad (4.1)$$

for all $x \subseteq X, y \subseteq Y, z \subseteq Z, \omega' \subseteq \Omega$, *where* $\mu_{pl^{\Omega}}$ *is associated with any* pl^{Ω} *in* $\mathscr{P}_{\mathrm{r}}^{\Omega}$.

Subsets ω' in Eq. (4.1) are said to be *indistinguishable* with respect to x on X, y on Y and z on Z (i.e., their respective projections on X, Y and Z coincide with each other). Intuitively, $\mu_{pl_{\mathrm{r}}^{\Omega}}(x,y,z)$ 'clamps' together $\mu_{pl^{\Omega}}(\omega')$ of all indistinguishable subsets ω' associated with (x,y,z) on Ω, considering only their total μ value rather than any specific μ distribution among ω'.

Lemma 4.2 pl_{r}^{Ω} *and* $\mu_{pl_{\mathrm{r}}^{\Omega}}$ *are one-to-one correspondent and related with each other through:* $\forall (x,y,z) \in \mathrm{r}_\Omega$,

$$pl_{\mathrm{r}}^{\Omega}(x,y,z) = \textstyle\sum_{x' \subseteq x, y' \subseteq y, z' \subseteq z} \mu_{pl_{\mathrm{r}}^{\Omega}}(x',y',z').$$

Proof Suppose there exists some $\omega' \subseteq \Omega$ such that $\mu_{pl^{\Omega}}(\omega')$ contributes to (i) more than one element in $\mu_{pl_{\mathrm{r}}^{\Omega}}$, the elements must be the same (by Def. 2); (ii) none of the elements in $\mu_{pl_{\mathrm{r}}^{\Omega}}$, ω' must be $\emptyset$ (by Def. 2). From this, $\mu_{pl_{\mathrm{r}}^{\Omega}}$ can be inferred as a lossless coarsening of $\mu_{pl^{\Omega}}$, and thus by Eq. (2.3), the above relation can be obtained.

Consequently, Lemma 4.1 can be equivalently written in terms of $\mu_{pl_{\mathrm{r}}^{\Omega}}$ (see Lemma 4.3), and so the initially defined relation between any node and its parent (see Lemma 4.4).

Lemma 4.3 *For any given* pl_{r}^{Ω} *defined on* Ω, *we have:*

$$\begin{aligned}
\mu_{pl_{\mathrm{r}}^{\Omega}}^{\downarrow X}(x) &= \textstyle\sum_{y \subseteq Y, z \subseteq Z} \mu_{pl_{\mathrm{r}}^{\Omega}}(x,y,z), \\
\mu_{pl_{\mathrm{r}}^{\Omega}}^{\downarrow XY}(x,y) &= \textstyle\sum_{z \subseteq Z} \mu_{pl_{\mathrm{r}}^{\Omega}}(x,y,z), \\
\mu_{pl_{\mathrm{r}}^{\Omega}}[z]^{\downarrow XY}(x,y) &= \textstyle\sum_{z' \subseteq z} \mu_{pl_{\mathrm{r}}^{\Omega}}(x,y,z'), \\
\mu_{pl_{\mathrm{r}}^{\Omega}}[y,z]^{\downarrow X}(x) &= \textstyle\sum_{y' \subseteq y, z' \subseteq z} \mu_{pl_{\mathrm{r}}^{\Omega}}(x,y',z'),
\end{aligned}$$

for all $x \subseteq X$, $y \subseteq Y$ *and* $z \subseteq Z$.

Lemma 4.4 *The initial relation between any* Y *and its parent* X *in Eq.* (2.6) *can be equivalently written by means of* $\mu_{pl_{\mathrm{r}}^{\Theta}}$ *on* $\Theta = XY$ *for all* $x \subseteq X$, $y \subseteq Y$ *and* $\alpha_x \triangleq (-1)^{|x|+1}$ *as*

$$\mu_{pl_{\mathrm{r}}^{\Theta}}(x,y) = \alpha_x \prod_{x_i \in x} pl_{\mathrm{r}}^{\Theta}(x_i,y) - \sum_{y' \subset y} \mu_{pl_{\mathrm{r}}^{\Theta}}(x,y'). \quad (4.2)$$

Proof Eq. (4.2) is obtained by first rewriting Eq. (2.6) as $pl_{\mathrm{r}}^{\Theta}(x,y) = 1 - \prod_{x_i \in x}(1 - pl_{\mathrm{r}}^{\Theta}(x_i,y))$ (4.3) using Lemma (4.1), then replacing all pl_{r} terms in Eq. (4.3) with their corresponding $\mu_{pl_{\mathrm{r}}}$ using Lemma 4.2.

With all quantitative aspects of a network represented in the form of pl_{r} (by means of its corresponding $\mu_{pl_{\mathrm{r}}}$), belief propagation is now essentially concerned with combination of such pl_{r}. For instance, pl^{AG} (Figure 2b) is the combined impact of pl_{r}^{A}, pl_{r}^{AB}, pl_{r}^{B} and pl_{r}^{BG} (Figure 2a) on AG. We will next show how pl_{r} can be efficiently and directly combined by means of $\mu_{pl_{\mathrm{r}}}$. Since the algorithm is based on the message-passing mechanism, belief combination methods will be presented from the perspective of each node.

4.2 COMBINATION OF pl_{r} FUNCTIONS

At any node X_0 in the network, belief combination using LC can be generalised into two scenarios: the set of pl_{r} to be combined are (1) mutually distinct (they do not share any common conditioning variables) and (2) non-distinct (they are otherwise). Across nodes, Case 1 is concerned with belief combination associated with 'tree' portions of loop(s) (e.g., combination of pl_{r}^{AG} and pl_{r}^{FG} at node G to propagate belief from A to F); while Case 2 with computing the entire impact of one or more loops on X_0 (e.g., combination of pl_{r}^{AG} and $pl_{\mathrm{r}}^{AG'}$ at node G). As such let $\mathbf{X} = \{X_0, ..., X_n\}$ be any subset of nodes in the network; $\mathbf{U}_1$ and $\mathbf{U}_2$ respectively be any set of subsets of $\mathbf{X}$ (without loss of generality, assume $\mathbf{X} = \mathbf{U}_1 \cup \mathbf{U}_2$) and $\mathbf{V} = \mathbf{U}_1 \cap \mathbf{U}_2$. Let Θ_1, Θ_2 and Ω be the Cartesian product of the elements in $\mathbf{U_1}$, $\mathbf{U_2}$, and $\mathbf{X}$, respectively, we are interested in the combined impact on Ω for any given $pl_{\mathrm{r}}^{\Theta_1}$ and $pl_{\mathrm{r}}^{\Theta_2}$ where $\mathbf{U}_1$ and $\mathbf{U}_2$ can be arbitrarily overlapping.

4.2.1 Combination of Distinct pl_{r} Functions

Theorem 4.5 is concerned with belief combination at X_0 in Case 1 where $pl_{\mathrm{r}}^{\Theta_1}$ and $pl_{\mathrm{r}}^{\Theta_2}$ to be combined are distinct (i.e., having non-overlapping conditioning variables, or $\mathbf{V} = \{X_0\}$). The theorem proposes a new conjunctive combination rule $\circledcap_r$ such that the combined impact pl_{r}^{Ω}=$pl_{\mathrm{r}}^{\Theta_1} \circledcap_r pl_{\mathrm{r}}^{\Theta_2}$ on Ω is equivalent to the partially defined pl_{r} associated with $pl^{\Omega} = pl^{\Theta_1} \circledcap pl^{\Theta_2}$ (for any $pl^{\Theta_1} \in \mathscr{P}_{\mathrm{r}}^{\Theta_1}$ and $pl^{\Theta_2} \in \mathscr{P}_{\mathrm{r}}^{\Theta_2}$) without having to perform the latter computations, the knowledge of which is generally not available, and the computations significantly more expensive.

Theorem 4.5 *Given $pl_{\mathrm{r}}^{\Theta_1}$ and $pl_{\mathrm{r}}^{\Theta_2}$ defined on Θ_1 and Θ_2, respectively where $\mathbf{V} = \{X_0\}$, their combined impact $pl_{\mathrm{r}}^{\Omega} = pl_{\mathrm{r}}^{\Theta_1} \circledcap_r pl_{\mathrm{r}}^{\Theta_2}$ on Ω is uniquely determined by means of their corresponding $\mu_{pl_{\mathrm{r}}^{\Omega}}$, $\mu_{pl_{\mathrm{r}}^{\Theta_1}}$ and $\mu_{pl_{\mathrm{r}}^{\Theta_2}}$:*

$$\mu_{pl_{\mathrm{r}}^{\Omega}}(\omega) = (-1)^{|x_0|+1} \mu_{pl_{\mathrm{r}}^{\Theta_1}}(\omega^{\downarrow\Theta_1}) \mu_{pl_{\mathrm{r}}^{\Theta_2}}(\omega^{\downarrow\Theta_2}), \quad (4.4)$$

for all $\omega \in \mathrm{r}_{\Omega}$ and $x_0 = \omega^{\downarrow X_0}$.
Proof See the Appendix.

When there are more than two mutually distinct pl_{r} to be combined at a node, Eq. (4.4) can be carried out in a pairwise fashion. When they are across nodes, application of $\circledcap_r$ in Theorem 4.5 and marginalisation in Lemma 4.1 and 4.3 fulfill the three axioms by Shenoy [Shenoy1993] that make local computations possible. For instance, let $pl_{\mathrm{r}1}, pl_{\mathrm{r}2}, pl_{\mathrm{r}3}$ and $pl_{\mathrm{r}12}$ denote pl_{r}^{AB}, pl_{r}^{BG}, pl_{r}^{FG} and pl_{r}^{ABG} in Figure 2, respectively, the following axioms are satisfied for all $a \subseteq A$, $b \subseteq B$, $f \subseteq F$, and $g \subseteq G$:

Commutativity and associativity of combination, e.g.,
$((\mu_{pl_{\mathrm{r}1}} \circledcap_r \mu_{pl_{\mathrm{r}2}}) \circledcap_r \mu_{pl_{\mathrm{r}3}})(a, b, f, g)$
$= (-1)^{|g|+1} \left((-1)^{|b|+1} \mu_{pl_{\mathrm{r}1}}(a, b) \mu_{pl_{\mathrm{r}2}}(b, g)\right) \mu_{pl_{\mathrm{r}3}}(f, g)$
$= (-1)^{|b|+1} \left((-1)^{|g|+1} \mu_{pl_{\mathrm{r}2}}(b, g) \mu_{pl_{\mathrm{r}3}}(f, g)\right) \mu_{pl_{\mathrm{r}1}}(a, b)$
$= (\mu_{pl_{\mathrm{r}1}} \circledcap_r (\mu_{pl_{\mathrm{r}2}} \circledcap_r \mu_{pl_{\mathrm{r}3}}))(a, b, f, g)$

Consonance of marginalisation, e.g,
$pl_{\mathrm{r}}^{ABG\downarrow AB\downarrow A}(a) = \sum_{b\subseteq B} (\sum_{g\subseteq G} \mu_{pl_{\mathrm{r}12}}(a, B, G))$
$= \sum_{b\subseteq B, g\subseteq G} \mu_{pl_{\mathrm{r}12}}(a, B, G) = pl_{\mathrm{r}}^{ABG\downarrow A}(a)$

Distributivity of marginalisation over combination, e.g.,
$(pl_{\mathrm{r}1} \circledcap_r pl_{\mathrm{r}2})^{\downarrow AB} = \sum_g (-1)^{|b|+1} \mu_{pl_{\mathrm{r}1}}(a, b) \mu_{pl_{\mathrm{r}2}}(b, g)$
$= (-1)^{|b|+1} \mu_{pl_{\mathrm{r}1}}(a, b) \sum_g \mu_{pl_{\mathrm{r}2}}(b, g) = pl_{\mathrm{r}1} \circledcap_r (pl_{\mathrm{r}2})^{\downarrow B}$.

When $pl_{\mathrm{r}}^{\Theta_1}$ and $pl_{\mathrm{r}}^{\Theta_2}$ are distinct, pl_{r}^{Ω} can be exactly and uniquely determined due to the cylindrical extension $(\omega^{\downarrow\Theta_1})^{\uparrow\Omega}$ (resp. $(\omega^{\downarrow\Theta_2})^{\uparrow\Omega}$) applied to $\omega^{\downarrow\Theta_1}$ (resp. $\omega^{\downarrow\Theta_2}$) for each $\omega \in \mathrm{r}_{\Omega}$, rendering the product in Eq. (4.4) insensitive to any specific distribution of μ among those *indistinguishable* ω' (see Def. (2)) associated with each $\omega \in \mathrm{r}_{\Omega}$.

4.2.2 Combination of Non-distinct pl_{r} Functions

When $pl_{\mathrm{r}}^{\Theta_1}$ and $pl_{\mathrm{r}}^{\Theta_2}$ are non-distinct (i.e., having common conditioning variables: $|\mathbf{V}|>1$ in Case 2), the equality in Eq. (4.4) no longer holds, thus their combination result on Ω is no longer unique. The basic mathematical reasoning for Theorem 4.6 that determines a set of possible pl_{r}^{Ω} resulting from combination of $pl_{\mathrm{r}}^{\Theta_1}$ and $pl_{\mathrm{r}}^{\Theta_2}$ is as follows. Let

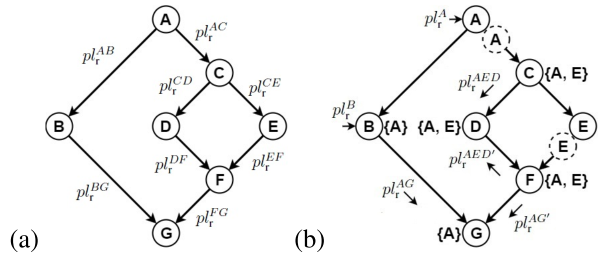


Figure 2: (a) A sample network where the relations between nodes are represented with pl_{r}, and (b) with its loops broken by the cutset $\{A, E\}$, and example messages received by D and G during belief updating.

- $\Phi_i = \Omega \backslash_{X_i}$, $\Phi_{ij} = \Omega \backslash_{\{X_i, X_j\}}$ denote the product of elements in $\mathbf{X} \backslash_{X_i}$ and $\mathbf{X} \backslash_{\{X_i, X_j\}}$, respectively;
- $\omega \in \mathrm{r}_{\Omega}$ be equivalently represented as $\omega = (x_i, \phi_i)$ where $x_i = \omega^{\downarrow X_i}$, $\phi_i = \omega^{\downarrow\Phi_i}$ for any $X_i \in \mathbf{X}$[3], and as $\omega = (x_i, x_j, \phi_{ij})$ for any $X_j (\neq X_i) \in \mathbf{X}$, $x_j = \omega^{\downarrow X_j}$ and $\phi_{ij} = \omega^{\downarrow\Phi_{ij}}$;
- $N(\omega)$ denote the number of $X_i \in \mathbf{V}$ associated with ω s.t. $x_i = \omega^{\downarrow X_i}$ is a non-singleton subset (i.e., $|x_i| > 1$)[4];
- α_{x_i} denote $(-1)^{|x_i|+1}$ for some $x_i \subseteq X_i$; and
- ω be arranged in partial order by increasing size $N(\omega)$.

We are now interested in computing $pl_{\mathrm{r}}^{\Omega}(\omega)$ for all $\omega \in \mathrm{r}_{\Omega}$.

(I) We first consider all those ω with $N(\omega) \leq 1$, i.e., there exists $X_i \in \mathbf{V}$ s.t. $\omega = (x_i, \phi_i)$ and $N(\phi_i) = 0$. Since $pl_{\mathrm{r}}^{\Omega}(x_i, \phi_i) = pl_{\mathrm{r}}^{\Omega}[\phi_i]^{\downarrow X_i}(x_i)$ for all $x_i \subseteq X_i$ (Lemma 4.1) and $N(\phi_i) = N(\phi_i^{\downarrow\Theta_1}) = N(\phi_i^{\downarrow\Theta_2}) = 0$ in this case, only *singleton* subsets of the conditioning variables associated with both $pl_{\mathrm{r}}^{\Theta_1}$ and $pl_{\mathrm{r}}^{\Theta_2}$ hold, thereby eliminating the associated loop(s) and rendering the belief functions induced by $pl_{\mathrm{r}}^{\Theta_1}$ and $pl_{\mathrm{r}}^{\Theta_2}$ on X_i, $pl_{\mathrm{r}}^{\Theta_1}[\phi_i^{\downarrow\Theta_1}]^{\downarrow X_i}$ and $pl_{\mathrm{r}}^{\Theta_2}[\phi_i^{\downarrow\Theta_2}]^{\downarrow X_i}$, distinct. Thus by Theorem 4.5 we have for all $x_i \subseteq X_i$:

$$\mu_{pl_{\mathrm{r}}^{\Omega}}(x_i, \phi_i) = \alpha_{x_i} \mu_{pl_{\mathrm{r}}^{\Theta_1}}(x_i, \phi_i^{\downarrow\Theta_1}) \mu_{pl_{\mathrm{r}}^{\Theta_2}}(x_i, \phi_i^{\downarrow\Theta_2}),$$
$$pl_{\mathrm{r}}^{\Omega}(x_i, \phi_i) = \sum_{x_{i'} \subseteq x_i, \phi_{i'} \subseteq \phi_i, \phi_{i'} \in \mathrm{r}_{\Phi_i}} \mu_{pl_{\mathrm{r}}^{\Omega}}(x_{i'}, \phi_{i'}). \quad (4.5)$$

All $pl_{\mathrm{r}}^{\Omega}(\omega)$ computed by Eq. (4.5) are referred to collectively as the *Core* of pl_{r}^{Ω} ($\mathscr{C}_{pl_{\mathrm{r}}^{\Omega}}$) from which the bounds associated with $pl_{\mathrm{r}}^{\Omega}(\omega)$ where $N(\omega) \geq 2$ are derived below.

(II) We next consider all ω where $N(\omega) = 2$, i.e., there exist X_i and X_j in $\mathbf{V}$ s.t. $\omega = (x_i, x_j, \phi_{ij})$, $|x_i| > 1$, $|x_j| > 1$. With respect to X_i alone, we have by Lemma 4.2:

$$pl_{\mathrm{r}}^{\Omega}(\omega) = pl_{\mathrm{r}}^{\Omega}(x_i, \phi_i) = \sum_{x_{i_k}} pl_{\mathrm{r}}^{\Omega}(x_{i_k}, \phi_i) + \sum_{x_{i*}, \phi_{i'}} \mu_{pl_{\mathrm{r}}^{\Omega}}(x_{i*}, \phi_{i'}), \quad (4.6)$$

for all $x_{i_k} \in x_i$, $x_{i*} (\neq x_{i_k}) \subseteq x_i$, $\phi_{i'} (\in \mathrm{r}_{\Phi_i}) \subseteq \phi_i$.

Regarding the first sum of Eq. (4.6), since $(x_{i_k}, \phi_i) = (x_{i_k}, x_j, \phi_{ij})$ and $N(x_{i_k}, \phi_{ij}) = 0$, $pl_{\mathrm{r}}^{\Omega}(x_{i_k}, \phi_i)\ \forall x_{i_k} \in x_i$ can be exactly computed using Eq. (4.5). Regarding the second sum, $\mu_{x_{i*}, \phi_{i'}} \triangleq \sum_{x_{i*}, \phi_{i'}} \mu_{pl_{\mathrm{r}}^{\Omega}}(x_{i*}, \phi_{i'})$, $\mu_{x_{i*}, \phi_{i'}}$ can be considered the *total overlapping (belief) mass* and thus constrained by those $pl_{\mathrm{r}}^{\Omega}(x_{i_k}, \phi_i)\ \forall x_{i_k} \in x_i$. Due to

[3] For instance, $\omega = (x_0, ..., x_n)$ can be represented as $\omega = (x_1, \phi_1)$ where $\phi_1 = (x_0, x_2, ..., x_n)$ with respect to X_1.

[4] For instance, given $\omega = (x_0, ..., x_n)$ and $\mathbf{V} = \{X_0, X_1\}$, $N(\omega) = 0$ when $|x_0| = |x_1| = 1$, $N(\omega) = 1$ when $|x_0| > 1$ or $|x_1| > 1$, and $N(\omega) = 2$ when $|x_0| > 1$ and $|x_1| > 1$.

$\alpha_{x_{i*}} \sum_{\phi_{i'}} \mu_{pl_r^\Omega}(x_{i*}, \phi_{i'}) \leq \min_{x_{ik*} \in x_{i*}} pl_r^\Omega(x_{x_{ik*}, \phi_i})$[5] substitution of which into $\mu_{x_{i*}, \phi_{i'}}$ of Eq. (4.6) yields:

$$\underline{pl_r^\Omega}(\omega) = \underline{pl_r^\Omega}(x_i, \phi_i) = \max_{x_{i_k}} \; pl_r^\Omega(x_{i_k}, \phi_i) \quad (4.7)$$

$$\overline{pl_r^\Omega}(\omega) = \overline{pl_r^\Omega}(x_i, \phi_i) = \min(1, \textstyle\sum_{x_{i_k}} pl_r^\Omega(x_{i_k}, \phi_i)) \quad (4.8)$$

corresponding to Fréchet bounds in probability theory. Since x_j is also non-singleton, $\mu_{x_{i*}, \phi_{i'}}$ is expected to be further constrained, e.g., through $pl_r^\Omega(x_{j_l}, \phi_j), \forall x_{j_l} \in x_j$.

Let pl_r^Θ denote $pl_r^\Omega[\phi_{ij}]^{\Downarrow\Theta}$, $\Theta{=}X_iX_j$, we can decompose $pl_r^\Omega(\omega)$ with respect to X_j in reference to X_i as

$$\begin{aligned} & pl_r^\Omega(\omega) = pl_r^\Omega(x_i, x_j, \phi_{ij}) = pl_r^\Omega[\phi_{ij}]^{\Downarrow\Theta}(x_i, x_j) \\ &= \sum_{x_{j_l}} pl_r^\Omega(x_i, x_{j_l}, \phi_{ij}) \quad + \sum_{x_{i'}, x_{j*}, \phi_{ij'}} \mu_{pl_r^\Omega}(x_{i'}, x_{j*}, \phi_{ij'}) \\ &= \sum_{x_{j_l}} pl_r^\Omega[\phi_{ij}]^{\downarrow\Theta}(x_i, x_{j_l}) \; + \sum_{x_{i'}, x_{j*}} \mu_{pl_r^\Omega}[\phi_{ij}]^{\downarrow\Theta}(x_{i'}, x_{j*}) \\ &= \sum_{x_{j_l}} pl_r^\Theta(x_i, x_{j_l}) \quad + \sum_{x_{i'}, x_{j*}} \mu_{pl_r^\Theta}(x_{i'}, x_{j*}), \end{aligned} \quad (4.9)$$

for all $x_{i'} {\subseteq} x_i$, $x_{j_l} {\in} x_j$, $x_{j*}({\neq} x_{j_l}) {\subseteq} x_j$, $\phi_{ij'}({\in}\, \mathrm{r}_{\Phi_{ij}}) {\subseteq} \phi_{ij}$; and subsequently $pl_r^\Theta(x_i, x_{j_l})$ with respect to X_i as

$$pl_r^\Theta(x_i, x_{j_l}) = \textstyle\sum_{x_{i_k}} pl_r^\Theta(x_{i_k}, x_{j_l}) + \sum_{x_{i*}} \mu_{pl_r^\Theta}(x_{i*}, x_{j_l}) \quad (4.10)$$

for all $x_{i_k} {\in} x_i$ and $x_{i*}({\neq} x_{i_k}) \subseteq x_i$.

From Eqs. (4.6) and (4.10), we have

$$\begin{aligned} \mu_{x_{i*}, \phi_{i'}} &\leq \min_{x_{j_l}} \; \textstyle\sum_{x_{i*}} \mu_{pl_r^\Theta}(x_{i*}, x_{j_l})^6 \\ &\leq \min_{x_{j_l}} (pl_r^\Theta(x_i, x_{j_l}) - \textstyle\sum_{x_{i_k}} pl_r^\Theta(x_{i_k}, x_{j_l})), \end{aligned}$$

thus the bound in Eq. (4.8) can be further tightened:

$$\begin{aligned} \overline{pl_r^\Omega}(\omega) = \min\bigl(1, &\textstyle\sum_{x_{i_k}} pl_r^\Omega(x_{i_k}, \phi_i) + \\ &\min_{x_{j_l} \in x_j} (pl_r^\Theta(x_i, x_{j_l}) - \textstyle\sum_{x_{i_k}} pl_r^\Theta(x_{i_k}, x_{j_l}))\bigr) \end{aligned} \quad (4.11).$$

(III) Finally, we consider all ω where $N(\omega){>}2$. Each ω in this case is associated with multiple such X_i in Eq. (4.6), each in turn associated with multiple such X_j in Eq. (4.9), this necessitates extending Eqs (4.7) and (4.11) to all $X_i \in \mathbf{X}$, and all $X_j({\neq} X_i) \in \mathbf{X}$, respectively. Since $N(x_{i_k}, \phi_i) = N(x_i, \phi_i) - 1$, Eqs (4.7) and (4.11) can be efficiently applied in a dynamic programming manner to all ω with increasing size $N(\omega)$ as shown in Theorem 4.6.

Theorem 4.6 *Given any two non-distinct $pl_r^{\Theta_1}$ and $pl_r^{\Theta_2}$ defined on Θ_1 and Θ_2, respectively (i.e., $|\mathbf{V}| > 1$), their combined impact on Ω, pl_r^Ω, can be approximated as:*

Step 1: Computation of $\mathscr{C}_{pl_r^\Omega}$:

$$\mu_{pl_r^\Omega}(x_i, \phi_i) {=} \alpha_{x_i} \mu_{pl_r^{\Theta_1}}(x_i, \phi_i^{\downarrow\Theta_1}) \mu_{pl_r^{\Theta_2}}(x_i, \phi_i^{\downarrow\Theta_2})$$

$$pl_r^\Omega(x_i, \phi_i) \; = \textstyle\sum_{x_{i'} \subseteq x_i, \phi_{i'} (\in \mathrm{r}_{\Phi_i}) \subseteq \phi_i} \mu_{pl_r^\Omega}(x_{i'}, \phi_{i'}) \quad (4.12)$$

for all $\omega = (x_i, \phi_i) \in \mathrm{r}_\Omega$ where $|x_i| \geq 1$ and $N(\phi_i) = 0$.

Step 2: Reconstruction of pl_r^Ω:

$$\underline{pl_r^\Omega}(\omega) = \max_{X_i} (\max_{x_{i_k} \in x_i} \underline{pl_r^\Omega}(x_{i_k}, \phi_i)) \quad (4.13)$$

$$\overline{pl_r^\Omega}(\omega) = \min_{X_i} (1, \sum_{x_{i_k} \in x_i} \overline{pl_r^\Omega}(x_{i_k}, \phi_i) + \min_{X_j} \mu_{x_{i*}, \phi_{i'}}) \quad (4.14)$$

$$\mu_{x_{i*}, \phi_{i'}} \triangleq \min_{x_{j_l} \in x_j} (\overline{pl_r^\Omega}(x_i, x_{j_l}, \phi_{ij}) - \sum_{x_{i_k}} \overline{pl_r^\Omega}(x_{i_k}, x_{j_l}, \phi_{ij}))$$

for all ω s.t. $N(\omega) > 1$, all $X_i, X_j {\in} \mathbf{X}$ s.t. $|x_i|{>}1$, $|x_j|{>}1$.

Since $pl_r^{\Omega\downarrow X_i}(x_i) = pl_r^\Omega(x_i, \Phi_i)$, Theorem 4.6 allows simultaneous approximation of the posterior marginal of any node X_i in the form of a set of possible belief functions defined by the bounds imposed on $pl_r^{\Omega\downarrow X_i}(x_i), \forall x_i \subseteq X_i$. It is highly desirable however in many cases to approximate a single combination result (which would be more informative, and more efficient with respect to belief propagation). As such, we will now take one step further and devise a method that estimates a specific point value $pl_r^\Omega(\omega)$ within the interval approximation $[\underline{pl_r^\Omega}(\omega), \overline{pl_r^\Omega}(\omega)]$ in Theorem 4.6 in such a way that $pl_r^\Omega(\omega)$ *reflects the current state of knowledge* with a high degree of accuracy.

4.2.3 Approximation with parameterised DRC/GBT

Recall from Lemma 4.1 that $pl_r^\Omega(x_i, \phi_i) = pl_r^\Omega[\phi_i]^{\downarrow X_i}(x_i)$, $\forall x_i \subseteq X_i$. By Eq. (4.12), a 'complete' marginal of pl_r^Ω on some X_i (i.e., $pl^\Omega[\phi_i]^{\downarrow X_i} {\equiv} pl_r^\Omega[\phi_i]^{\downarrow X_i}$) can be exactly computed when ϕ_i holds, rendering $\mathscr{C}_{pl_r^\Omega}$ a collective set of such marginals on all $X_i \in X$. As a result, Theorem 4.6 Step 2 can be reformulated as the problem of reconstructing pl_r^Ω from those marginals. To this end, with respect to X_i and X_j in Eq. (4.9), let us rewrite the equation in the form of conditional belief functions on X_i in reference to X_j:

For all $x_{i_k} {\in} x_i$, $x_{i*}({\neq} x_{i_k}) {\subseteq} x_i$: $pl_r^{X_i}[x_j] {\triangleq} pl_r^\Omega[x_j, \phi_{ij}]^{\downarrow X_i}$,

$$\begin{aligned} pl_r^\Omega(\omega) &= pl_r^{X_i}[x_j](x_i) \\ &= \sum_{x_{i_k}} pl_r^{X_i}[x_j](x_{i_k}) + \sum_{x_{i*}} \mu_{pl_r^{X_i}}[x_j](x_{i*}), \end{aligned} \quad (4.15)$$

where $pl_r^{X_i}[x_j](x_{i_k})$, $\forall x_{i_k} {\in} x_i$ have been computed in the previous iteration. Applying DRC/GBT as in Eq. (2.6) which directly computes $pl_r^{X_i}[x_j](x_i)$ from $pl_r^{X_i}[x_j](x_{i_k})$ is not eligible since (i) those $pl_r^{X_i}[x_j](x_{i_k})$ are not the only knowledge about $pl_r^\Omega(\omega)$, and (ii) they are not initially defined but correspond to pl_r^Ω on Ω and thus generally not MCP-compatible. This requires us to extend the *Generalised Likelihood Principle* (GLP) [Smets1994] to a multivariate context without the MCP-compatibility assumption.

Extended GLP For any $pl_r^\Omega(\omega)$ in Eq. (4.15), we have:

(1) $pl_r^\Omega(\omega) = pl_r^\Omega[x_j, \phi_{ij}]^{\downarrow X_i}(x_i) = pl_r^\Omega[x_i, \phi_{ij}]^{\downarrow X_j}(x_j)$
$= pl_r^{X_i}[x_j](x_i) = pl_r^{X_j}[x_i](x_j)$ for all $X_i, X_j {\in} \mathbf{X}$

(2) $pl_r^{X_i}[x_j](x_i)$ is related to $pl_r^{X_i}[x_j](x_{i_k})$, $\forall x_{i_k} \in x_i$ for any x_i and x_j through a parameterised function $\mathscr{F}^{s_{[x_j]x_i}}$ where $s_{[x_j]x_i}$ is informed by knowledge of $pl_r^{X_k}[x_i](x_{k_t}) {\triangleq} pl_r^\Omega[x_i, \phi_{ik}]^{\downarrow X_k}(x_{k_t})$, $\forall x_{k_t} \in x_k$, for all $X_k({\neq} X_i) \in \mathbf{X}$[7].

The extended GLP allows $pl_r^\Omega(\omega)$ to be computed in the form of $pl_r^{X_i}[x_j](x_i)$ for any X_i and X_j through an estimation of the parameter $s_{[x_j]x_i}$. In this regard, estimation of $s_{[x_j]x_i}$ for each $x_j {\subseteq} X_j$ is motivated by the dual aspect of $pl_r^{\Theta = X_iX_j}$, the conditional form of which is given in Eq. (4.15). With respect to Ω, pl_r^Θ is a *marginalisation* of

[5] As $(-1)^{|A|+1}\mu(A) \geq (-1)^{|B|+1}\mu(B{\supseteq}A)$ for any subset A and B, and $\mu_{pl}(A) \equiv pl(A)$ when $|A| = 1$.

[6] $\mu_{x_{i*}\phi_{i'}} = \sum_{x_{i*} x_{j'}} \mu_{pl_r^\Omega}[\phi_{ij}]^{\Downarrow\Theta}(x_{i*}, x_{j'}) = \sum_{x_{i*} x_{j'},} \mu_{pl_r^\Theta}(x_{i*}, x_{j'})$.

[7] Note that all those $pl_r^{X_i}[x_j](x_{i_k})$ and $pl_r^{X_k}[x_i](x_{k_t})$ were already computed in the previous iteration.

pl_{r}^{Ω} on Θ (through conditioning on ϕ_{ij}: $pl_{\mathrm{r}}^{\Theta} = pl_{\mathrm{r}}^{\Omega}[\phi_{ij}]^{\downarrow\Theta}$). With respect to Θ, it can be considered a *specialisation* [Klawonn and Smets1992] of $pl_{\mathrm{r}}^{\Theta_0}$ initially defined on Θ which is MCP-compatible[8], thus satisfying DRC/GBT:

$$pl_{\mathrm{r}}^{X_i 0}[x_{j'}](x_i) = 1 - \prod_{x_{i_k} \in x_i} (1 - pl_{\mathrm{r}}^{X_i 0}[x_{j'}](x_{i_k})) \quad (4.16)$$

for all $x_{j'} \subseteq x_j$. Such $pl_{\mathrm{r}}^{X_i 0}[x_{j'}](x_i)$ corresponds to a point within the interval $[\underline{pl_{\mathrm{r}}^{X_i 0}}[x_{j'}](x_i),\ \overline{pl_{\mathrm{r}}^{X_i 0}}[x_{j'}](x_i)]$ in Theorem 4.6 Step 2. As $pl_{\mathrm{r}}^{\Theta_0}$ is specialised into pl_{r}^{Θ}, the equality in Eq. (4.16) may no longer hold, moving the corresponding $pl_{\mathrm{r}}^{X_i}[x_{j'}](x_i)$ for each $x_{j'} \subseteq x_j$ toward either of its respective bounds. To this end, let $r_{[x_j]x_i} \in [-1, 1]$[9] be associated with $pl_{\mathrm{r}}^{X_i}[x_j](x_i)$ to linearly measure the relative mass overlapping between $pl_{\mathrm{r}}^{X_i}[x_j](x_{i_k}), \forall x_{i_k} \in x_i$ such that $r_{[x_j]x_i} = 0, 1, -1$ when $pl_{\mathrm{r}}^{X_i}[x_j](x_i)$ coincides with its MCP-compatible value (determined by Eq. (4.16)), and its lower bound and upper bound (determined by Eqs. (4.13) and (4.14)), respectively. Then for each $x_j \subseteq X_j$, a parameterised version of DRC/GBT can be defined as the Frank t-conorm [Klement, Mesiar, and Pap2013], $\mathscr{F}^{s_{[x_j]x_i}}$, for any other arbitrary value of $r_{[x_j]x_i}$:

$$pl_{\mathrm{r}}^{X_i}[x_j](x_i) = \mathscr{F}^{s}_{x_{i_k}} (pl_{\mathrm{r}}^{X_i}[x_j](x_{i_k}))$$

$$= 1 - \log_s \left(1 + \frac{\prod_{x_{i_k}} \left(s^{1 - pl_{\mathrm{r}}^{X_i}[x_j](x_{i_k})} - 1 \right)}{(s-1)^{n-1}} \right) \quad (4.17)$$

where $s \triangleq s_{[x_j]x_i}$ is a positive parameter defined in terms of $r_{[x_j]x_i}$ as $s_{[x_j]x_i} = \pi(1 - r_{[x_j]x_i})/4$ [Ferson et al.2004].

This alternative formulation, in place of Eq. (4.15), presents an interesting property when MCP-compatibility holds:

$$r^0_{[x_j]x_i} = r^0_{[x_{j_l}]x_i} = 0 \text{ for all } x_{j_l} \in x_j, \quad (4.18)$$

where $r^0_{[x_{j_l}]x_i}$ are associated with $pl_{\mathrm{r}}^{X_i 0}[x_{j_l}](x_i), \forall x_{j_l} \in x_j$. As $pl_{\mathrm{r}}^{\Theta 0}$ is specialised into pl_{r}^{Θ}, and accordingly $r^0_{[x_{j_l}]x_i}$ to $r_{[x_{j_l}]x_i}$, $\forall x_{j_l} \in x_j$, we are interested in computing $r_{[x_j]x_i}$ from those $r_{[x_{j_l}]x_i}$ whose knowledge is available from the previous iteration. We conjecture among all the arbitrary $pl_{\mathrm{r}}^{\Theta +}$ that satisfies the constraints imposed by such $r_{[x_{j_l}]x_i}$, the one that exhibits the least divergence from the initial relation in Eq. (4.18) would likely correspond to pl_{r}^{Θ}, e.g.,

$$r_{[x_j]x_i} = \arg\min_{r^+_{[x_j]x_i}} \sum_{x_{j_l}} pl_{\mathrm{r}}^{X_i}[x_{j_l}](x_i)(r^+_{[x_j]x_i} - r_{[x_{j_l}]x_i})^2,$$

allowing $r_{[x_j]x_i}$ to be approximated as a weighted average:

$$r_{[x_j]x_i} = \sum_{x_{j_l}} w_{[x_{j_l}]x_i} r_{[x_{j_l}]x_i},\ w_{[x_{j_l}]x_i} = \frac{pl_{\mathrm{r}}^{X_i}[x_{j_l}](x_i)}{\sum_{x_{j_l}} pl_{\mathrm{r}}^{X_i}[x_{j_l}](x_i)}. \quad (4.19)$$

Utilising parameterised t-norms and t-conorms for combining belief functions induced by possibly overlapping bodies of evidence has been studied thoroughly in [Denœux2008] where the relative dependence between the belief functions is (i) defined by a single parameter s (i.e., $s = s_{[x_j]x_i} = s_{[x_{j'}]x_i}, \forall x_{j'} \subset x_j$ in this case) which can be learned from data, and (ii) concerned with positive dependence (i.e., $s \in [0, 1]$). In contrast, our method deals with situations where the parameters $s_{[x_{j_l}]x_i}$ are (i) generally of different values computed during execution time, and (ii) concerned with the entire range of positive and negative dependence (i.e., $s_{[x_{j'}]x_i} \in [0, +\infty), \forall x_{j'} \subseteq x_j$).

By replacing Eqs. (4.13) and (4.14) with (4.17) and (4.19), the interval approximation for each $pl_{\mathrm{r}}^{\Omega}(\omega)$ in Theorem 4.6 Step 2 can be reduced to a point estimation[10] (below), where the pair X_i and X_j in Eq. (4.20) correspond to those in $\mathbf{X}$ such that $pl_{\mathrm{r}}^{\Omega}[\phi_{ij}]^{\downarrow\Theta}$ exhibits the least divergence from its initial MCP-compatible value $pl_{\mathrm{r}}^{\Theta 0}$ on $\Theta = X_i X_j$.

Step 2: Reconstruction of pl_{r}^{Ω}:

$$pl_{\mathrm{r}}^{X_i}[x_j](x_i) \triangleq pl_{\mathrm{r}}^{\Omega}[x_j, \phi_{ij}]^{\downarrow X_i}(x_i)$$

$$pl_{\mathrm{r}}^{\Omega}(\omega) = pl_{\mathrm{r}}^{X_i}[x_j](x_i) = \mathscr{F}^{s_{[x_j]x_i}}_{x_{ik} \in x_i}(pl_{\mathrm{r}}^{X_i}[x_j](x_{i_k})), \quad (4.20)$$

where $s_{[x_j]x_i} = \pi(1 - r_{[x_j]x_i})/4$, $r_{[x_j]x_i} = \min_{X_p} \min_{X_q} |r_{[x_q]x_p}|$ $\forall X_p, X_q \in \mathbf{X}$ s.t. $x_p = \omega^{\downarrow X_p}$, $x_q = \omega^{\downarrow X_q}$ and $|x_p| > 1$, $|x_q| > 1$.

Since both the interval approximation in Theorem 4.6 Step 2, and point estimation in Eq. (4.20), of any $pl_{\mathrm{r}}^{\Omega}(\omega), \omega \in \mathrm{r}_{\Omega}$ is reconstructed based on $\mathscr{C}_{pl_{\mathrm{r}}}$, it is sufficient that beliefs be propagated in the form of $\mathscr{C}_{pl_{\mathrm{r}}}$[11], hence Fact 4.7.

Fact 4.7 *Given a node X_0 associated with $(n-1)$ conditioning variables, all with m states, the messages propagated to X_0 in the form of $\mathscr{C}_{pl_{\mathrm{r}}}$ have space complexity of $\mathcal{O}(2^{mm} - Z)$ where $Z = \sum_{k=2}^{n} \binom{n}{k} m^{n-k}(2^m - m)^k$.*

5 DISCUSSION AND RESULTS

By means of local conditioning (LC), the proposed method enables belief propagation directly on the original multiply-connected network. The advantages of maintaining and propagating beliefs in the original structure of the network are multi-fold, e.g., eliminating the assumption/requirement that the reasoning network is singly-connected at and during runtime, mitigating disruptions during execution associated with generation of a secondary structure due to changes in the network topology, facilitating local computations in a distributed manner and allowing run-time pruning of the network, and enhancing transparency of reasoning and supporting explanation.

By combining local conditioning and conditional belief functions, the proposed method allows combination and propagation of beliefs with a significant reduction in space

[8] By construction, this is the case when X_i and X_j are directly connected; otherwise, they are assumed to have an implicit edge with vacuous cbfs which are naturally MCP-compatible.

[9] Let pl^0, $\underline{pl}$ and $\overline{pl}$ denote the MCP-compatible value, lower and upper bounds associated with the pl to be computed, respectively, the relative mass overlapping r can be defined as follows:

$$r = \begin{cases} (pl - pl^0)/(pl^0 - \overline{pl}), & \overline{pl} \geq pl \geq pl^0 \\ (pl - pl^0/(\underline{pl} - pl^0), & pl^0 > pl \geq \underline{pl} \\ \text{undefined}, & \text{otherwise.} \end{cases}$$

[10] Note that this estimation is not guaranteed to produce a valid belief function, thus a normalisation step may be required.

[11] For instance, in Figure 2b, the beliefs propagated to G would be $\mathscr{C}_{pl_{\mathrm{r}}^{AG}}$ and $\mathscr{C}_{pl_{\mathrm{r}}^{AG'}}$ rather than pl_{r}^{AG} and $pl_{\mathrm{r}}^{AG'}$.

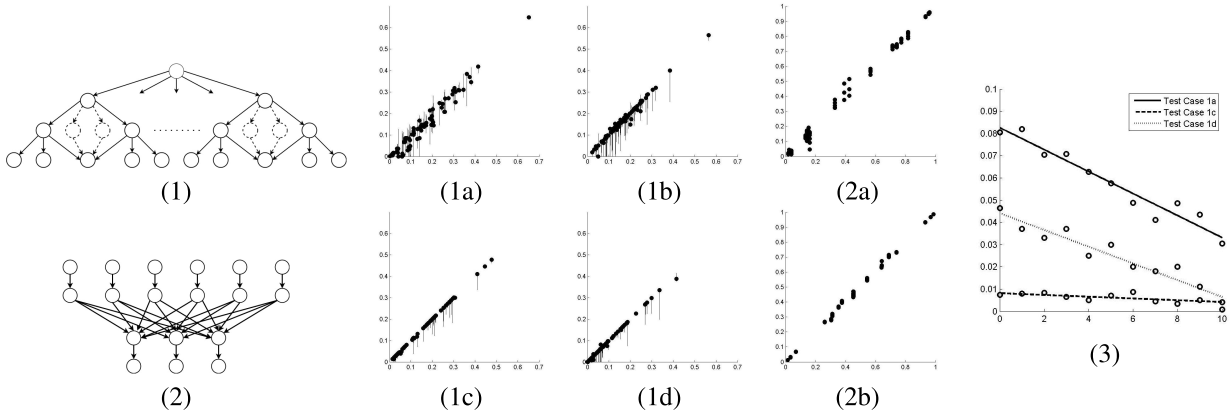


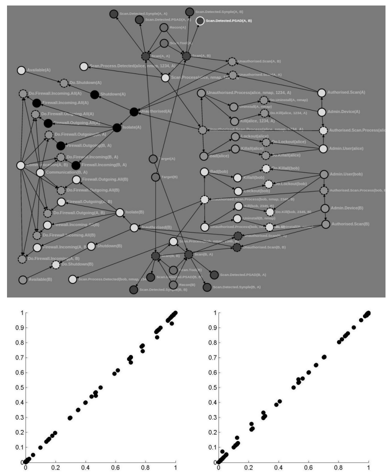
Figure 3: (1) A sample hierarchical network with 10 clusters of loops, and the correlated posterior marginals (normalised beliefs) for Test Case (1a), (1b), (1c) and (1d). (2) A sample densely connected network and the correlated posterior marginals for (2a) a worst case scenario i.e., no observations, and (2b) where three random nodes in the network receive (soft) observations. (3) Average approximation error for Test Case 1a, 1c and 1d as a function of observations received.

and time. As shown in Fact 4.7, the worst case space complexity associated with the messages propagated in the form of $\mathscr{C}_{pl_r}$ is only a fraction of the associated conditional space ($\mathcal{O}(2^{mn})$), and a very small fraction of the product space ($\mathcal{O}(2^{m^n})$) if exact belief updating is to be performed.

One of the rationales for representing beliefs in the form of partially defined pl and $b.pl.a$ functions is to reduce the space complexity of the messages propagated, and thus the computational complexity involved. Reducing the size of belief functions for the purpose of efficient combination has received substantial attention in the literature, resulting in a rich class of belief transformation methods (see [Bauer1997], [Wilson2000], [Cuzzolin2012]). To this end, one could potentially argue that instead of representing the beliefs in the form of pl and $b.pl.a$ functions as proposed, they could be defined on the product space with a smaller number of focal sets by means of belief transformation. However the unsuitability of such alternative approaches can be readily identified at a high level. Specifically, they would suffer from: (i) a significant computational overhead involved with belief transformation during propagation, (ii) information loss due to beliefs being forced to be more 'committed' during propagation, and (iii) inconsistent results due to violation of the axioms that make local computations possible since most belief transforms do not commute with combination. In contrast, our proposed technique does not experience the above problems, specifically (i) no computational overhead since it enables simultaneous combination operating directly on the cbfs initially defined between nodes without requiring any belief transformation, (ii) mitigated information loss: instead of transforming the correct belief function to be propagated into an approximate one with a smaller number of focal elements, it propagates the *Core*, $\mathscr{C}_{pl_r}$, of the 'correct'[12] beliefs on the product space, facilitating reconstruction of the original belief functions on demand, and (iii) consistency of results is also significantly improved as belief combination and propagation between nodes with the same conditioning variables satisfies the axioms facilitating local computations. In another vein, stochastic sampling [Wilson2000, Laâmari, Hariz, and Yaghlane2014] could also be applied to combine beliefs received at a node. However, the methods aim to reduce the computational complexity associated with belief combination, but not to reduce the space complexity involved with each message sent and received.

As mentioned in Section 3, unlike LBP which is known for not producing the correct numerical value for the posterior marginals of nodes, our proposed method is aimed at approximating the correct posterior marginals for the purpose of belief updating. Thus it is important to verify the degree of accuracy of the results produced by the approach. To this end, experiments have been performed to verify the practical accuracy of our approximate algorithm. Due to space limitations, we will primarily present representative cases. In Test Cases 1, we conducted experiments on networks which are instantiations of Figure 3.1. We first started with networks with minimal information and complexity (Test Case 1a), e.g., each branch of the network contains only the simplest form of loop (i.e., dashed-line nodes are excluded), each node is associated with a binary domain, prior beliefs are vacuous, and none of the nodes receive any observations during belief updating. We then introduced more information and complexity to the networks, specifically, a (soft) observation was introduced at every loop in Test Case (1b), the complexity of loops in each branch was increased (i.e., dashed-line nodes included) in Test Case (1c), and the domain size of each node was increased (to tertiary domain) in Test Case (1d). Figures 3.1a, 1b, 1c and 1d present the posterior marginals approximated for every node in Test Case 1a, 1b, 1c and 1d, respectively, correlated with the correct results

[12] Provided that the beliefs propagated to nodes outside the respective loops are correct.



(a) (b) (c)

Figure 4: Correlated posterior marginals in the form of belief (b) and plausibility (c) functions obtained for a real-world (CND) application (a).

generated using the exact method based on LC and joint belief function briefly proposed in Section 3. To illustrate each approximate result with respect to its lower and upper bounds, we also approximated the lower and upper posterior marginals for each node, according to Theorem 4.6, depicted as a vertical bar associated with each marginal. As the figures illustrate, the algorithm produced approximate results with a high degree of accuracy. Even in the worst case scenarios where all the prior beliefs and observations are totally vacuous (Test Case 1a), the approximate marginals are close to exact; and with some increase in the 'informativeness' and complexity of the network, the bounds obtained for Test Case 1b, 1c and 1d were tightened and the approximate marginals are almost exact. Simulating the networks with different sets of randomly generated parameters produces similar results. Figure 3.3 shows the average error (vertical axis) of the approximate marginals — compared to exact, using L_2 norm [Cuzzolin2012] — obtained from 100 randomly generated instances each of Test Case 1a, 1c and 1d, each with 0 to 10 observations (horizontal axis). The results show that on average, the approximation error across all the tests is low even in the absence of any observations, and diminishes with additional observations.

To reveal the behaviour of the algorithm when dealing with more extreme loops and arbitrary networks, we extended our experiments with the densely-connected network in Figure 3.2 and a real-world network drawn from our CND application (Figure 4a). For both cases, the algorithm produced accurate results. Indeed, in cases where the network contains complex loops such as these, $\mathscr{C}_{pl_{\mathrm{r}}}$ tends to become a substantial belief dominant region of pl_{r}. Not only does this help increase the accuracy of the results, but also the approximation process quickly converges, significantly reducing the space and computational complexity involved.

6 CONCLUSIONS

This paper proposed an approximate message-passing belief propagation method for multiply-connected networks. By means of local conditioning, the algorithm does not require the network to be singly-connected prior to execution. By having knowledge of the network encoded and belief combination formulated by means of the proposed partially defined pl and $b.pl.a$ functions, local conditioning can be combined with (subsets of) conditional belief functions, approximating belief updating in an efficient manner. The empirical results demonstrated that the method is able to produce results with a high degree of accuracy while achieving a significant reduction in space and computational complexity (in comparison to its exact counterpart), promoting the practicality of networks with (conditional) belief functions. Indeed, the algorithm is currently being incorporated into Influx, the inference engine of our CND system, in order to address the aforementioned challenges that motivated this research effort.

Acknowledgement The author thanks Michael Docking for providing advice and discussion throughout this work.

A APPENDIX

Proof (Theorem 4.5) Let $\Phi_1 = \Theta_1 \backslash X_0$, $\Phi_2 = \Theta_1 \backslash X_0$[13], and $\alpha_{x_0} \triangleq (-1)^{|x_0|+1}$. Since the variables associated with Φ_1 and Φ_2 in this case are disjoint, the pieces of evidence induced on X_0 when $\phi_1 \in \mathrm{r}_{\Phi_1}$ and $\phi_2 \in \mathrm{r}_{\Phi_2}$ hold are distinct. Thus the plausibility that ϕ_1 and ϕ_2 jointly induce on X_0 can be obtained by Lemma 2.1, that is:

$$\mu_{pl_{\mathrm{r}}^{\Omega}}[\phi_1, \phi_2]^{\downarrow X_0}(x_0) = \alpha_{x_0} \mu_{pl_{\mathrm{r}}^{\Theta_1}}[\phi_1]^{\downarrow X_0}(x_0) \mu_{pl_{\mathrm{r}}^{\Theta_2}}[\phi_2]^{\downarrow X_0}(x_0), \forall x_0 \subseteq X, \quad \text{(A.1)}$$

which can be equivalently written by Lemma 4.3 as:

$$\sum_{\phi_{1'}, \phi_{2'}} \mu_{pl_{\mathrm{r}}^{\Omega}}(x_0, \phi_{1'}, \phi_{2'}) = \alpha_{x_0} \sum_{\phi_{1'}} \mu_{pl_{\mathrm{r}}^{\Theta_1}}(x_0, \phi_{1'}) \sum_{\phi_{2'}} \mu_{pl_{\mathrm{r}}^{\Theta_2}}(x_0, \phi_{2'}), \forall \phi_{i'} \subseteq \phi_i, \phi_{i'} \in \mathrm{r}_{\Phi_i}, i = 1, 2. \quad \text{(A.2)}$$

Let us investigate the impact that ϕ_1 and ϕ_2 (with increasing cardinalities) jointly induce on X_0. By Eq. (A.2), we have for all $\omega^* = (x_0, \phi_{1^*}, \phi_{2^*})$ s.t. $|\phi_{1^*}| = 1$ and $|\phi_{2^*}| = 1$:

$$\mu_{pl_{\mathrm{r}}^{\Omega}}(x_0, \phi_{1^*}, \phi_{2^*}) = \alpha_{x_0} \mu_{pl_{\mathrm{r}}^{\Theta_1}}(x_0, \phi_{1^*}) \mu_{pl_{\mathrm{r}}^{\Theta_2}}(x_0, \phi_{2^*}). \quad \text{(A.3)}$$

For all $\omega' = (x_0, \phi_{1'}, \phi_{2^*})$ s.t. $|\phi_{1'}| = 2$ and $|\phi_{2^*}| = 1$:

$$\mu_{pl_{\mathrm{r}}^{\Omega}}(x_0, \phi_{1'}, \phi_{2^*}) = \sum_{\phi_{1''} \subseteq \phi_{1'}} \mu_{pl_{\mathrm{r}}^{\Omega}}(x_0, \phi_{1''}, \phi_{2^*}) = \alpha_{x_0} \sum_{\phi_{1''} \subseteq \phi_{1'}} \mu_{pl_{\mathrm{r}}^{\Theta_1}}(x_0, \phi_{1''}) \mu_{pl_{\mathrm{r}}^{\Theta_2}}(x_0, \phi_{2^*}). \quad \text{(A.4)}$$

Substitute terms in the LHS of Eq. (A.4) with terms in the RHS of Eq. (A.3) and simplify the result:

$$\mu_{pl_{\mathrm{r}}^{\Omega}}(x_0, \phi_{1'}, \phi_{2^*}) = \alpha_{x_0} \mu_{pl_{\mathrm{r}}^{\Theta_1}}(x_0, \phi_{1'}) \mu_{pl_{\mathrm{r}}^{\Theta_2}}(x_0, \phi_{2^*}). \quad \text{(A.5)}$$

For each $x_0 \subseteq X_0$, iteratively applying the above steps to all $\phi_1 \in \mathrm{r}_{\Phi_1}$, $\phi_2 \in \mathrm{r}_{\Phi_2}$ with increasing cardinalities, we have:

$$\mu_{pl_{\mathrm{r}}^{\Omega}}(x_0, \phi_1, \phi_2) = \alpha_{x_0} \mu_{pl_{\mathrm{r}}^{\Theta_1}}(x_0, \phi_1) \mu_{pl_{\mathrm{r}}^{\Theta_2}}(x_0, \phi_2).$$

Thus for all $\omega \in \mathrm{r}_{\Omega}$ and $x_0 = \omega^{\downarrow X_0}$:

$$\mu_{pl_{\mathrm{r}}^{\Omega}}(\omega) = (-1)^{|x_0|+1} \mu_{pl_{\mathrm{r}}^{\Theta_1}}(\omega^{\downarrow \Theta_1}) \mu_{pl_{\mathrm{r}}^{\Theta_2}}(\omega^{\downarrow \Theta_2}).$$

[13] Φ_1, Φ_2 denotes the product of the variables associated with Θ_1, Θ_2, respectively, excluding X_0.

References

[Bauer1997] Bauer, M. 1997. Approximation algorithms and decision making in the dempster-shafer theory of evidence – an empirical study. *International Journal of approximate reasoning* 17(2):217–237.

[Cuzzolin2012] Cuzzolin, F. 2012. The geometry of uncertainty. *Information Science and Statistics Series. Springer*.

[Denœux2008] Denœux, T. 2008. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence* 172(2-3):234–264.

[Díez1996] Díez, F. J. 1996. Local conditioning in bayesian networks. *Artificial Intelligence* 87(1):1–20.

[Dubois and Prade2009] Dubois, D., and Prade, H. 2009. Formal representations of uncertainty. *Decision-Making Process: Concepts and Methods* 85–156.

[Fay and Jaffray2000] Fay, A., and Jaffray, J.-Y. 2000. A justification of local conditioning in bayesian networks. *International Journal of Approximate Reasoning* 24(1):59–81.

[Ferson et al.2004] Ferson, S.; Nelsen, R. B.; Hajagos, J.; Berleant, D. J.; Zhang, J.; Tucker, W. T.; Ginzburg, L. R.; and Oberkampf, W. L. 2004. Dependence in probabilistic modeling, dempster-shafer theory, and probability bounds analysis. *Sandia National Laboratories, Report No. SAND2004-3072*.

[Klawonn and Smets1992] Klawonn, F., and Smets, P. 1992. The dynamic of belief in the transferable belief model and specialization-generalization matrices. In *Proceedings of the Eighth international conference on Uncertainty in artificial intelligence*, 130–137. Morgan Kaufmann Publishers Inc.

[Klement, Mesiar, and Pap2013] Klement, E. P.; Mesiar, R.; and Pap, E. 2013. *Triangular norms*, volume 8. Springer Science & Business Media.

[Laâmari, Hariz, and Yaghlane2014] Laâmari, W.; Hariz, N. B.; and Yaghlane, B. B. 2014. Approximate inference in directed evidential networks with conditional belief functions using the monte carlo algorithm. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, 486–497. Springer.

[Pearl1988] Pearl, J. 1988. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann.

[Shafer, Shenoy, and Mellouli1987] Shafer, G.; Shenoy, P. P.; and Mellouli, K. 1987. Propagating belief functions in qualitative markov trees. *International Journal of Approximate Reasoning* 1(4):349–400.

[Shafer1976] Shafer, G. 1976. *A mathematical theory of evidence*, volume 1. Princeton university press Princeton.

[Shenoy and Shafer1986] Shenoy, P. P., and Shafer, G. 1986. Propagating belief functions with local computations. *IEEE Expert* 1(3):43–52.

[Shenoy1993] Shenoy, P. P. 1993. Valuation networks and conditional independence. In *Proceedings of the Ninth international conference on Uncertainty in artificial intelligence*, 191–199. Morgan Kaufmann Publishers Inc.

[Shenoy1997] Shenoy, P. P. 1997. Binary join trees for computing marginals in the shenoy-shafer architecture. *International Journal of approximate reasoning* 17(2):239–263.

[Smets and Kennes1994] Smets, P., and Kennes, R. 1994. The transferable belief model. *Artificial intelligence* 66(2):191–234.

[Smets1994] Smets, P. 1994. Belief functions: the disjunctive rule of combination and the generalized bayesian theorem. *International Journal of approximate reasoning* 9(1):1–35.

[Weiss1997] Weiss, Y. 1997. Belief propagation and revision in networks with loops.

[Wilson2000] Wilson, N. 2000. Algorithms for dempster-shafer theory. In *Handbook of defeasible reasoning and uncertainty management systems*. Springer. 421–475.

[Xu and Smets1996] Xu, H., and Smets, P. 1996. Reasoning in evidential networks with conditional belief functions. *International Journal of Approximate Reasoning* 14(2):155–185.

[Yaghlane and Mellouli2008] Yaghlane, B. B., and Mellouli, K. 2008. Inference in directed evidential networks based on the transferable belief model. *International Journal of Approximate Reasoning* 48(2):399–418.

[Yaghlane, Smets, and Mellouli2002a] Yaghlane, B. B.; Smets, P.; and Mellouli, K. 2002a. Belief function independence: I. the marginal case. *International Journal of Approximate Reasoning* 29(1):47–70.

[Yaghlane, Smets, and Mellouli2002b] Yaghlane, B. B.; Smets, P.; and Mellouli, K. 2002b. Belief function independence: Ii. the conditional case. *International Journal of Approximate Reasoning* 31(1):31–75.