Efficient Computation of Belief Theoretic Conditionals

Lalintha G. Polpitiya, Kamal Premaratne, Manohar N. Murthi and Dilip Sarkar

University of Miami
Coral Gables, Florida (USA)

10th International Symposium on Imprecise Probability: Theories and Applications (ISIPTA '17)
Lugano, Switzerland, 10-14 July 2017

Acknowledgement: This work is based on research supported by the U.S. Office of Naval Research (ONR) via grant #N00014-10-1-0140, and the U.S. National Science Foundation (NSF) via grant #1343430.

12 July 2017

Outline

Motivation and Challenges

- **Dempster-Shafer (DS)** theory offers greater expressiveness and flexibility in evidential reasoning.
- **Conditional operation:** plays a pivotal role in DST strategies for evidence updating and fusion.
- **Implementations:** restricted to smaller frames of discernment (FoDs) because of the prohibitive computational burden that larger FoDs impose on existing methods.
- **Several approximation methods:** compromise the quality of the generated results for computational efficiency.
- **Exact computation of conditionals is of paramount importance:** quality of results generated from DS theoretic (DST) strategies depend directly on the precision of the conditional.

Contributions

to fill the void between what DS theory can offer and its practical implementation

- The main contribution: a generalized computational model **(DS-Conditional-One)** for computing DST conditionals.

- This model can be employed to compute both the Fagin-Halpern (FH) and Dempster's conditional beliefs of an **arbitrary** proposition.

- This is exactly the challenge that Shafer refers to in [Shafer, 1990, p.348], viz., **"It remains to be seen how useful the fast Möbius transform will be in practice. It is clear, however, that it is not enough to make arbitrary belief function computations feasible."**

Contributions

to fill the void between what DS theory can offer and its practical implementation

- This new model can also be utilized as a **visualization tool** for conditional computations and in analyzing characteristics of conditioning operations.

- An outcome of this research is a **conditional computation library** which is available online.
(`https://profuselab.github.io/Conditional-Computation-Library/`)

Basic notions of belief theory

Symbol	Meaning
Θ	*Frame of discernment (FoD),* i.e., the set of all possible mutually exclusive and exhaustive propositions.
θ_i	*Singletons,* i.e., the lowest level of discernible information, i.e., $\Theta = \{\theta_0, \ldots, \theta_{n-1}\}$, here $n = \lvert\Theta\rvert$. For computational ease, we start the indexing from 0.
$\overline{A}$	*Complement* of the proposition $A \subseteq \Theta$, i.e., those singletons that are not in A.
$m(\cdot)$	*Basic belief assignment (BBA)* or *mass assignment* $m : 2^{\Theta} \mapsto [0, 1]$ where $\sum_{A \subseteq \Theta} m(A) = 1$ and $m(\emptyset) = 0$.
Focal element	Singleton or composite (i.e., non-singleton) proposition that receives a non-zero mass.
$\mathfrak{F}$	*Core,* the set of focal elements.
$\mathcal{E}$	*Body of evidence (BoE)* represented via the triplet $\{\Theta, \mathfrak{F}, m\}$.
Subset propositions of A	Subsets of A and A itself, here $m = \lvert A\rvert$

Straddling masses

The following notation is useful for our work:

Straddling masses

$$\mathcal{S}(A; B) = \sum_{\substack{\emptyset \neq C \subseteq A; \\ \emptyset \neq D \subseteq B}} m(C \cup D).$$

$\mathcal{S}(A; B)$ denotes the sum of all masses of propositions that 'straddle' both $A \subseteq \Theta$ and $B \subseteq \Theta$.

Γ_A and Π_A

The following result is of critical importance for our work:

Proposition 1

Consider the body of evidence (BoE) $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\cdot)\}$ and $A \subseteq \Theta$. For $B \subseteq \Theta$, consider the mappings $\Gamma_A : 2^\Theta \mapsto [0, 1]$ and $\Pi_A : 2^\Theta \mapsto [0, 1]$, where

$$\Gamma_A(B) = \sum_{\emptyset \neq X \subseteq \overline{A}} m((A \cap B) \cup X); \;\; \Pi_A(B) = \sum_{Y \subseteq (A \cap B)} \Gamma_A(Y).$$

Then the following are true:
(i) $\Gamma_A(A \cap B) = \Gamma_A(B)$ and $\Pi_A(A \cap B) = \Pi_A(B)$. So, w.l.o.g., we assume that $B \subseteq A$.
(ii) $\Gamma_A(\emptyset) = Bl(\overline{A})$.

Fagin-Halpern (FH) conditional

Most natural generalization of the probabilistic conditional notion - close connection with the inner and outer conditional probability measures.

Definition 2 (Fagin-Halpern (FH) conditional) [Fagin and Halpern, 1990]

Consider the BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\cdot)\}$ and $A \in \widehat{\mathfrak{F}}$. The *conditional belief* $Bl(B|A)$ of B given the conditioning event A is

$$Bl(B|A) = \frac{Bl(A \cap B)}{Bl(A \cap B) + Pl(A \cap \overline{B})}.$$ ■

Proposition 3

Consider the BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\cdot)\}$ and $A \in \widehat{\mathfrak{F}}$. Then, we may express $Bl(B|A)$ as

$$Bl(B|A) = \frac{Bl(A \cap B)}{1 - Bl(\overline{A}) - \mathcal{S}(\overline{A}; A \cap B)}, \quad B \subseteq \Theta.$$ □

Dempster's conditional

Perhaps the most widely employed DST conditional notion.

Definition 4 (Dempster's conditional) [Shafer, 1976]

Consider the BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\cdot)\}$ and $A \subseteq \Theta$ s.t. $Bl(\overline{A}) \neq 1$, or equivalently, $Pl(A) \neq 0$. The *conditional belief* $Bl(B \| A)$ of B given the conditioning event A is

$$Bl(B \| A) = \frac{Bl(\overline{A} \cup B) - Bl(\overline{A})}{1 - Bl(\overline{A})}. \quad \blacksquare$$

Proposition 5

Consider the BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\cdot)\}$ and $A \subseteq \Theta$ s.t. $Bl(\overline{A}) \neq 1$. Then, $Bl(B \| A)$ can be expressed as

$$Bl(B \| A) = \frac{Bl(A \cap B) + \mathcal{S}(\overline{A}; A \cap B)}{1 - Bl(\overline{A})}, \; B \subseteq \Theta. \quad \square$$

REcursive Generation of and Access to Propositions

REGAP: Starting with $\{\emptyset\}$ element [Polpitiya et al., 2016]

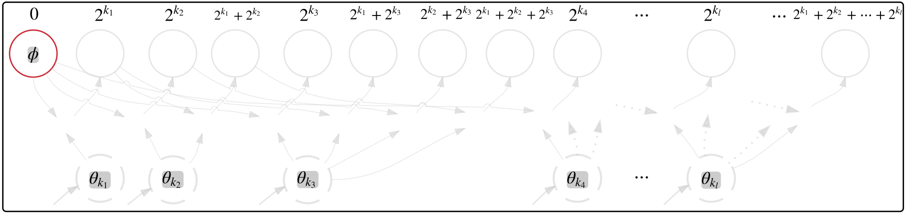

Figure: REGAP: *REcursive Generation of and Access to Propositions, Start with* $\emptyset$

- Consider the FoD $\Theta = \{\theta_0, \theta_1, \ldots, \theta_{n-1}\}$.
- Suppose we desire to determine the belief potential associated with $A = \{\theta_{k_1}, \theta_{k_2}, \ldots, \theta_{k_\ell}\} \subseteq \Theta$.
- The REGAP property allows us to recursively generate the propositions that are relevant for this computation: Start with $\{\emptyset\}$.

REcursive Generation of and Access to Propositions

REGAP: Inserting singleton $\{\theta_{k_3}\}$

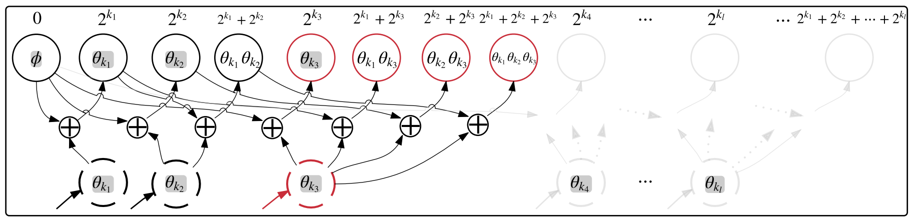

Figure: REGAP: *REcursive Generation of and Access to Propositions, inserting* $\{\theta_{k_3}\}$

- Inserting another singleton $\{\theta_{k_3}\} \in A$ brings the new propositions $\{\emptyset\} \cup \{\theta_{k_3}\} = \{\theta_{k_3}\}$, $\{\theta_{k_1}\} \cup \{\theta_{k_3}\} = \{\theta_{k_1}, \theta_{k_3}\}$, $\{\theta_{k_2}\} \cup \{\theta_{k_3}\} = \{\theta_{k_2}, \theta_{k_3}\}$, and $\{\theta_{k_1}, \theta_{k_2}\} \cup \{\theta_{k_3}\} = \{\theta_{k_1}, \theta_{k_2}, \theta_{k_3}\}$.
- In essence, when a new singleton is added, new propositions associated with it can be recursively generated by adding the new singleton to each existing proposition.

REcursive Generation of and Access to Propositions

REGAP: Generalized representation

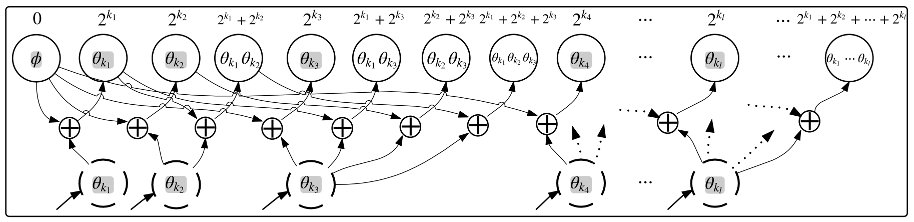

Figure: REGAP: *REcursive Generation of and Access to Propositions*

- *REGAP*, stands for *REcursive Generation of and Access to Propositions*.
- All propositions of interest within the FoD Θ can be generated when $A = \Theta$.
- These recursively generated propositions can be formulated as a vector, a matrix or a tree, and utilized to represent a dynamic BoE.

DS-Conditional-One computational model

- **DS-Conditional-One** is a computational model that enables one to compute the FH and Dempster's conditional beliefs of an **arbitrary** proposition.

- We denote the conditioning proposition A, its complement $\overline{A}$, and the conditioned proposition B as $\{a_0, a_1, \ldots, a_{|A|-1}\}$, $\{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $\{b_0, b_1, \ldots, b_{|B|-1}\}$, respectively. Here, $\Theta = \{\theta_0, \theta_1, \ldots, \theta_{n-1}\}$ denotes the FoD and $a_i, \alpha_j, b_k \in \Theta$.

- We represent singletons of the conditioning event $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ as *column singletons* and singletons of the complement of conditioning event $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$ as *row singletons* in a DS-Matrix.

DS-Conditional-One computational model

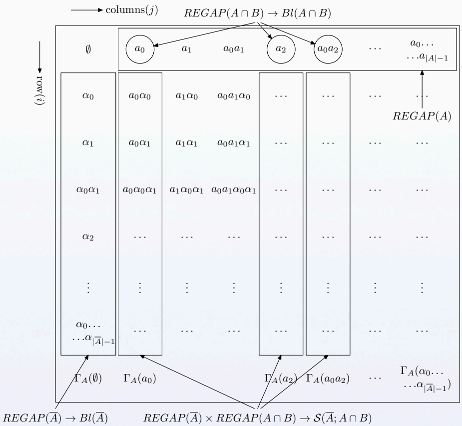

Figure: **DS-Conditional-One model.** Quantities related to $Bl(B|A)$ computation when $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ and $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $B = \{a_0, a_2\} \subseteq A$.

DS-Conditional-One computational model

- The proposed DS-Conditional-One computational model allows direct identification of $REGAP(A)$, $REGAP(\overline{A})$, $REGAP(A \cap B)$, $(REGAP(\overline{A}) \times REGAP(A \cap B))$, $(REGAP(A) \times REGAP(\overline{A}))$, and $\Gamma_A(C), \forall C \subseteq B$.

- We use a lookup table named *power* to enhance the computational efficiency. It contains 2 to the power of singleton indexes in increasing order.

- *index*[] is a dynamic array which keeps the indexes of subset propositions of $A \cap B$.

Compute $BI(A \cap B)$ (with time complexity $\mathcal{O}(2^{|A \cap B|})$)

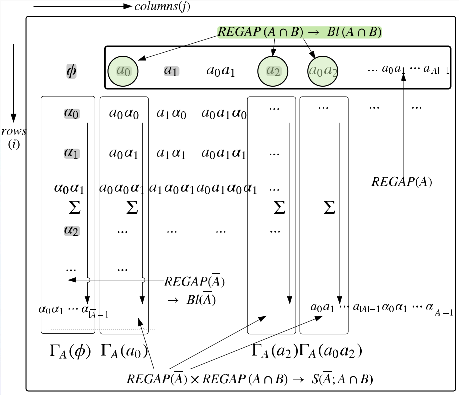

Figure: **DS-Conditional-One model.** Quantities related to $BI(B|A)$ computation when $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ and $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $B = \{a_0, a_2\} \subseteq A$.

Compute $BI(A \cap B)$: Algorithm 1

```
procedure BLB(Singletons A, Singletons B, DS-Matrix BBA)
    belief ← 0
    count ← 0
    for each a_i in A ∩ B do
        index[count] ← power[i]
        temp ← count
        count ← count + 1
        for j ← 0, temp − 1 do
            index[count] ← index[j] + power[i]
            count ← count + 1
        end for
    end for
    for i ← 0, power[|A ∩ B|] − 2 do
        belief ← belief + BBA[0][index[i]]
    end for
    Return belief
end procedure
```

Compute $BI(\overline{A})$ (with time complexity $\mathcal{O}(2^{|\overline{A}|})$)

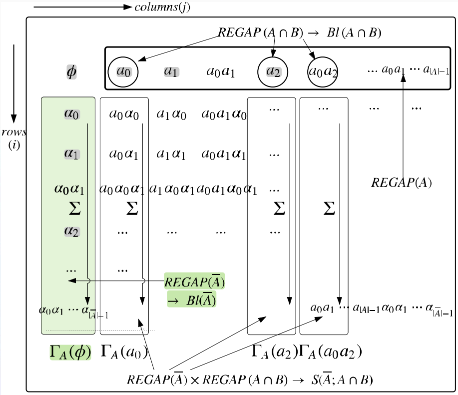

Figure: **DS-Conditional-One model.** Quantities related to $BI(B|A)$ computation when $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ and $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $B = \{a_0, a_2\} \subseteq A$.

Compute $Bl(\overline{A})$: Algorithm 2

procedure BLCOMP(Singletons $\overline{A}$, DS-Matrix BBA)

  $belief \leftarrow 0$

  for $i \leftarrow 1, power[|\overline{A}|] - 1$ **do**

    $belief \leftarrow belief + BBA[i][0]$

  end for

  Return *belief*

end procedure

Compute $\mathcal{S}(\overline{A}; A \cap B)$ (with time complexity $\mathcal{O}(2^{|\overline{A}|+|A\cap B|})$)

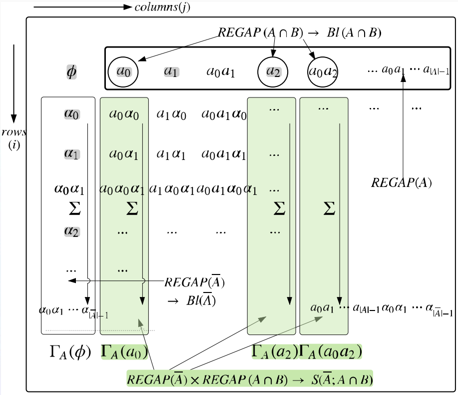

Figure: **DS-Conditional-One model.** Quantities related to $BI(B|A)$ computation when $A = \{a_0, a_1, \ldots, a_{|A|-1}\}$ and $\overline{A} = \{\alpha_0, \alpha_1, \ldots, \alpha_{|\overline{A}|-1}\}$, and $B = \{a_0, a_2\} \subseteq A$.

Compute $\mathcal{S}(\overline{A}; A \cap B)$: Algorithm 3

procedure STRAD(Singletons $\overline{A}$, Singletons A, Singletons B, DS-Matrix BBA)
  $belief \leftarrow 0$
  $count \leftarrow 0$
  for each a_i in $A \cap B$ **do**
    $index[count] \leftarrow power[i]$
    $temp \leftarrow count$
    $count \leftarrow count + 1$
    for $j \leftarrow 0,\ temp - 1$ **do**
      $index[count] \leftarrow index[j] + power[i]$
      $count \leftarrow count + 1$
    end for
  end for
  for $i \leftarrow 1,\ power[|\overline{A}|] - 1$ **do**
    for $j \leftarrow 0,\ power[|A \cap B|] - 2$ **do**
      $belief \leftarrow belief + BBA[i][index[j]]$
    end for
  end for
  Return $belief$
end procedure

FH conditional belief of an **arbitrary** proposition

Use the expression in Proposition 3, where $Bl(A \cap B)$, $Bl(\overline{A})$ and $\mathcal{S}(\overline{A}; A \cap B)$ are obtained via Algorithms 1, 2, and 3, respectively.

Proposition 3

Consider the BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\cdot)\}$ and $A \in \widehat{\mathfrak{F}}$. Then, we may express $Bl(B|A)$ as

$$Bl(B|A) = \frac{Bl(A \cap B)}{1 - Bl(\overline{A}) - \mathcal{S}(\overline{A}; A \cap B)}, \quad B \subseteq \Theta. \qquad \square$$

The computational complexity is $\mathcal{O}(2^{|\overline{A}|+|A \cap B|})$.

$Bl(B|A)$ when $B = \{a_0, a_2\}$

$$Bl(B|A) = \frac{Bl(A \cap B)}{1 - \Gamma_A(\{\emptyset\}) - \Gamma_A(\{a_0\}) - \Gamma_A(\{a_2\}) - \Gamma_A(\{a_0, a_2\})}.$$

Dempster's conditional belief of an **arbitrary** proposition

Use the expression in Proposition 5, where $Bl(A \cap B)$, $Bl(\overline{A})$ and $\mathcal{S}(\overline{A}; A \cap B)$ are obtained via Algorithms 1, 2, and 3, respectively.

Proposition 5

Consider the BoE $\mathcal{E} = \{\Theta, \mathfrak{F}, m(\cdot)\}$ and $A \subseteq \Theta$ s.t. $Bl(\overline{A}) \neq 1$. Then, $Bl(B \| A)$ can be expressed as

$$Bl(B \| A) = \frac{Bl(A \cap B) + \mathcal{S}(\overline{A}; A \cap B)}{1 - Bl(\overline{A})}, \; B \subseteq \Theta. \qquad \square$$

The computational complexity is $\mathcal{O}(2^{|\overline{A}|+|A \cap B|})$.

$Bl(B \| A)$ when $B = \{a_0, a_2\}$

$$Bl(B \| A) = \frac{Bl(A \cap B) + \Gamma_A(\{a_0\}) + \Gamma_A(\{a_2\}) + \Gamma_A(\{a_0, a_2\})}{1 - \Gamma_A(\{\emptyset\})}.$$

Dempster's conditional mass using specialization matrix [Smets, 2002]

- It employs a $2^{|\Theta|} \times 2^{|\Theta|}$-sized stochastic matrix $\mathfrak{S}_A$ (with each entry '0' or '1') referred to as the **conditioning specialization matrix** and a $2^{|\Theta|} \times 1$-sized vector $m(\cdot)$ containing the focal elements.

- The computational and space complexity of the specialization matrix multiplication is $\mathcal{O}(2^{|\Theta|} \times 2^{|\Theta|})$, a prohibitive burden even for modest FoD sizes.

Size of the DS-Conditional-One matrix vs Specialization matrix

When FoD size is 4

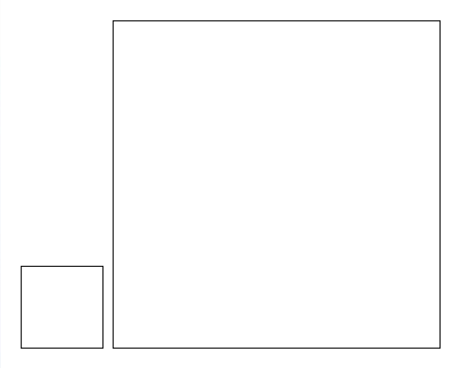

Figure: Size of the DS-Conditional-One matrix vs Specialization matrix: When FoD size is 4

Size of the DS-Conditional-One matrix vs Specialization matrix

When FoD size is 10

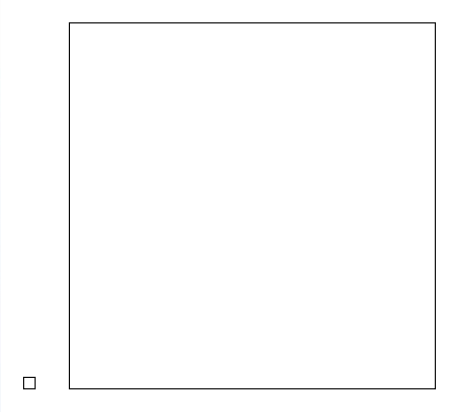

Figure: Size of the DS-Conditional-One matrix vs Specialization matrix: When FoD size is 10

Experiments

- For a given FoD size, we selected a random set of focal elements, with randomly selected mass values, and conducted 10,000 conditional computations for randomly chosen propositions A and $B \subseteq A$.

- With the DS-Conditional-One model (which applies to *both* FH and Dempster's conditionals), we use a 'brute force' approach to compute all the conditional beliefs. We then use the fast Möbius transform (FMT) [Thoma, 1989] to get the conditional masses for all the propositions.

- All conditional computations for an **arbitrary** proposition were done on an iMac running Mac OS X 10.12.3 (with 2.9GHz Intel Core i5 processor and 8GB of 1600MHz DDR3 RAM). Conditional computations for **all** propositions were done on the same iMac for smaller FoDs and on a supercomputer (`https://ccs.miami.edu/pegasus`) for larger FoDs (underlined in Table 1).

Experiments

Method → Conditional →		DS-Conditional-One Model FH or Dempster's			Specializa. Dempster's
FoD		$Bl(B\|A)$ or $Bl(B\|\|A)$	$Bl(B\|A)$ or $Bl(B\|\|A)$	$m(B\|A)$ or $m(B\|\|A)$	$m(B\|\|A)$
$\|\Theta\|$	Max. $\|\mathfrak{F}\|$	**(Arbitrary)**	**(All)**	**(All)**	**(All)**
2	3	0.0005	0.0011	0.0016	0.0011
4	15	0.0005	0.0038	0.0050	0.0063
6	63	0.0006	0.0128	0.0170	0.0696
8	255	0.0009	0.0517	0.0679	1.0154
10	1,023	0.0017	0.2428	0.3090	93.1590
12	4,095	0.0040	1.3528	1.6186	1,485.6300
14	16,383	0.0120	18.4885	22.4995	25,051.8200
16	65,535	0.0405	146.1480	151.9600	***
18	262,143	0.1516	1,087.2800	1,113.5300	***
20	1,048,575	0.6011	8,485.4500	8,862.9800	***

Table: DS-Conditional-One model versus specialization matrix based method. Average computational times (ms). (*** denotes computations not completed within a feasible time).

Concluding Remarks

- The significant speed advantage offered by the proposed computational model over the specialization matrix based approach is evident from experiments.
- For larger FoDs, the computational burden associated with the specialization matrix based approach becomes prohibitive because of its space complexity of $\mathcal{O}(2^{|\Theta|} \times 2^{|\Theta|})$. For example, an FoD of size 20 would need 128 $(= 2^{20} \times 2^{20}/8)$ GB of memory to represent the specialization matrix, if each matrix entry occupies only 1 bit.
- Our experiment results demonstrate that the average computational time taken to compute the conditional belief of an arbitrary proposition by the proposed approach is less than **2 (μs) for a FoD of size 10** and **0.7 (ms) for a FoD of size 20 ($\sim$1 million focal elements).**
- It also appears possible to further enhance the algorithms that we have developed via parallel computing optimizations because of the underlying matrix structure.

References I

Fagin, R. and Halpern, J. Y. (1990).
A new approach to updating beliefs.
In *Proc. 6th Conf. Uncertainty in Artificial Intelligence (UAI)*, pages 347–374, Cambridge, MA.

Polpitiya, L. G., Premaratne, K., Murthi, M. N., and Sarkar, D. (2016).
A framework for efficient computation of belief theoretic operations.
In *Proc. 19th Int. Conf. Information Fusion (FUSION)*, pages 1570–1577, Heidelberg, Germany.

Shafer, G. (1976).
A Mathematical Theory of Evidence.
Princeton Univ. Press, Princeton, NJ.

Shafer, G. (1990).
Perspectives on the theory and practice of belief functions.
Int. J. Approx. Reason., 4(5-6):323–362.

Smets, P. (2002).
The application of the matrix calculus to belief functions.
Int. J. Approx. Reason., 31(1):1–30.

Thoma, H. M. (1989).
Factorization of belief functions.
PhD thesis, Dept. Stat., Harvard Univ., Cambridge, MA.

Thank you very much!