Efficient Algorithms for Frequently Asked Questions

9. Solving Functional Aggregate Queries

Prof. Dan Olteanu

DaST

Data • (Systems+Theory)

University of Zurich UZH

May 16, 2022

https://lms.uzh.ch/url/RepositoryEntry/17185308706

Agenda for this Lecture

Solving Functional Aggregate Queries (FAQs)

- We start with FAQs over a single semiring, dubbed FAQ-SS
  - Uses LFTJ to compute the bags of a hypertree decomposition
  - Uses Yannakakis to aggregate away bound variables eagerly
- We then continue with FAQs over multiple semirings
  - Can we swap marginalisation order for variables under different semirings?
  - How can we deal with product aggregates?

Solving FAQ-SS

$$\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}} \cdots \bigoplus_{x_{n-1}} \bigoplus_{x_n} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$$

with hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ and semiring $(\mathbf{D}, \bigoplus, \bigotimes, \mathbf{0}, \mathbf{1})$

- Input: Hypertree decomposition $\mathcal{T}$ of Φ, factors $(\psi_S)_{S \in \mathcal{E}}$
- Runtime: $\mathcal{O}(N^{\text{fhtw}(\mathcal{T})} + \text{OUT})$ where OUT is the output size of Φ
- By choosing an optimal hypertree decomposition for the hypergraph $\mathcal{H}$ of Φ, the runtime becomes $\mathcal{O}(N^{\text{fhtw}(\mathcal{H})} + \text{OUT})$

Main Steps of the FAQ-SS Solver

Input: FAQ-SS Φ with hypergraph $\mathcal{H}$ and a hypertree decomposition $\mathcal{T}$ of $\mathcal{H}$

Step 1: Query Rewriting

1.1 Turn Φ into an equivalent α-acyclic FAQ Φ' by materialising the bags of $\mathcal{T}$

1.2 Choose an order σ for variable marginalisation *compatible* with $\mathcal{T}$

1.3 Rewrite Φ' into an FAQ Φ'' following σ by moving sums past products

Step 2: Query Evaluation

- Evaluate Φ'' by marginalising the bound variables *inside out*

We next explain the above steps using an example

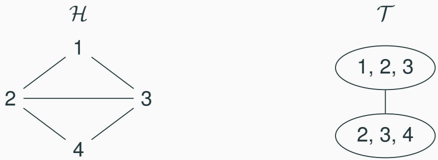
FAQ-SS Solver Example

Example of Solving FAQ-SS

Sum-product FAQ-SS:

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

with hypergraph $\mathcal{H}$ and hypertree decomposition $\mathcal{T}$:

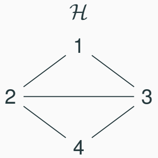


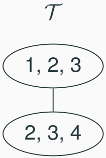
Example of Solving FAQ-SS

Sum-product FAQ-SS:

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

with hypergraph $\mathcal{H}$ and hypertree decomposition $\mathcal{T}$:





Φ can be computed in time $\mathcal{O}(N^2)$:

- Compute the join of the factors in time $\mathcal{O}(N^2)$ (since $\rho^*_{\mathcal{H}}(\{1, 2, 3, 4\}) = 2$)
- Marginalise out the variables X_2 and X_4

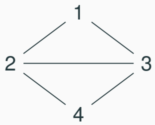
Example of Solving FAQ-SS

Sum-product FAQ-SS:

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

with hypergraph $\mathcal{H}$ and hypertree decomposition $\mathcal{T}$:

$\mathcal{H}$



$\mathcal{T}$

Φ can be computed in time $\mathcal{O}(N^2)$:

- Compute the join of the factors in time $\mathcal{O}(N^2)$ (since $\rho^*_{\mathcal{H}}(\{1, 2, 3, 4\}) = 2$)
- Marginalise out the variables X_2 and X_4

Next: a strategy to compute Φ in time $\mathcal{O}(N^{\frac{3}{2}})$

Step 1.1: Turning Φ into an α-acyclic FAQ

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

Turn Φ into an α-acyclic FAQ by materialising the bags of $\mathcal{T}$:

- Construct a factor $\psi_{123}(x_1, x_2, x_3)$ for B_1
- Construct a factor $\psi_{234}(x_2, x_3, x_4)$ for B_2
- Construct the α-acyclic query that has the join tree above:

$$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$$

Step 1.1: Turning Φ into an α-acyclic FAQ

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

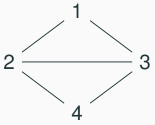
Turn Φ into an α-acyclic FAQ by materialising the bags of $\mathcal{T}$:

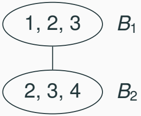
- Construct a factor $\psi_{123}(x_1, x_2, x_3)$ for B_1
- Construct a factor $\psi_{234}(x_2, x_3, x_4)$ for B_2
- Construct the α-acyclic query that has the join tree above:

$$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$$

How to construct ψ_{123} and ψ_{234}?

Constructing the Factors for the Bags (1/2)



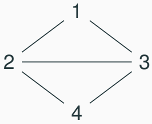


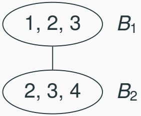
$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

First strategy: compute time

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3)$ $\mathcal{O}(N^2)$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$ $\mathcal{O}(N^{\frac{3}{2}})$

Constructing the Factors for the Bags (1/2)





$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

First strategy: compute time

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3)$ $\mathcal{O}(N^2)$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$ $\mathcal{O}(N^{\frac{3}{2}})$

Q: How can we reduce the compute time for ψ_{123}?

Constructing the Factors for the Bags (1/2)

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

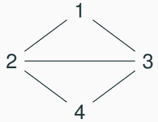
First strategy: compute time

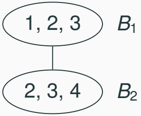
- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3)$ $\mathcal{O}(N^2)$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$ $\mathcal{O}(N^{\frac{3}{2}})$

Q: How can we reduce the compute time for ψ_{123}?

A: Include ψ_{23} into the computation of ψ_{123}

Constructing the Factors for the Bags (2/2)





$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

Second strategy: compute time

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3)$ $\mathcal{O}(N^{\frac{3}{2}})$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$ $\mathcal{O}(N^{\frac{3}{2}})$

Constructing the Factors for the Bags (2/2)

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

Second strategy: compute time

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3)$ $\mathcal{O}(N^{\frac{3}{2}})$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$ $\mathcal{O}(N^{\frac{3}{2}})$

$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3 x_4)$ is NOT equivalent to Φ

- The payloads of tuples in ψ_{23} are used twice in Φ'!

Constructing the Factors for the Bags (2/2)

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

Second strategy: compute time

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3)$ $\mathcal{O}(N^{\frac{3}{2}})$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$ $\mathcal{O}(N^{\frac{3}{2}})$

$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3 x_4)$ is NOT equivalent to Φ

- The payloads of tuples in ψ_{23} are used twice in Φ'!

Q: How to ensure Φ' is equivalent to Φ while keeping the compute time $\mathcal{O}(N^{\frac{3}{2}})$?

Constructing the Factors for the Bags (2/2)

$$\Phi(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

Second strategy: compute time

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3)$ $\mathcal{O}(N^{\frac{3}{2}})$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$ $\mathcal{O}(N^{\frac{3}{2}})$

$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3 x_4)$ is NOT equivalent to Φ

- The payloads of tuples in ψ_{23} are used twice in Φ'!

Q: How to ensure Φ' is equivalent to Φ while keeping the compute time $\mathcal{O}(N^{\frac{3}{2}})$?

A: Use the indicator projection of ψ_{23} in ψ_{234}

Indicator Projections

$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$, where

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3)$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$

In ψ_{234} we only need to know which tuples exist in ψ_{23}

Indicator Projections

$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$, where

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3)$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$

In ψ_{234} we only need to know which tuples exist in ψ_{23}

Indicator projection: New factor $\psi_{23/234}$ maps the tuples of ψ_{23} to 1:

$$\psi_{23/234}(x_2, x_3) = \begin{cases} \mathbf{1}, & \text{if } \psi_{23}(x_2, x_3) \neq \mathbf{0} \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

Notation $23/234$: $\psi_{23/234}$ only retains values for the variables in ψ_{234}

Indicator Projections

$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$, where

- $\psi_{123}(x_1, x_2, x_3) := \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3)$
- $\psi_{234}(x_2, x_3, x_4) := \psi_{23}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$

In ψ_{234} we only need to know which tuples exist in ψ_{23}

Indicator projection: New factor $\psi_{23/234}$ maps the tuples of ψ_{23} to 1:

$$\psi_{23/234}(x_2, x_3) = \begin{cases} \mathbf{1}, & \text{if } \psi_{23}(x_2, x_3) \neq \mathbf{0} \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

Notation $23/234$: $\psi_{23/234}$ only retains values for the variables in ψ_{234}

The factor for the second bag over variables $2, 3, 4$ is then:

$$\psi_{234}(x_2, x_3, x_4) := \psi_{23/234}(x_2, x_3) \cdot \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4)$$

Using Indicator Projections

Next strategy uses all indicator projections:

- Same asymptotic complexity as previous strategy: $\mathcal{O}(N^{\frac{3}{2}})$
- It may be more beneficial in practice

$$\psi_{123}(x_1, x_2, x_3) = \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{23}(x_2, x_3) \cdot \underbrace{\psi_{24/123}(x_2) \cdot \psi_{34/123}(x_3)}_{\text{indicator projections}}$$

$$\psi_{234}(x_2, x_3, x_4) = \psi_{24}(x_2, x_4) \cdot \psi_{34}(x_3, x_4) \cdot \underbrace{\psi_{12/234}(x_2) \cdot \psi_{13/234}(x_3) \cdot \psi_{23/234}(x_2, x_3)}_{\text{indicator projections}}$$

$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$ is equivalent to Φ

Step 1.2: Find Good Marginalisation Order and Rewrite Φ'

$$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$$

Step 1.2: Find Good Marginalisation Order and Rewrite Φ'

$$\Phi'(x_1, x_3) = \sum_{x_2} \sum_{x_4} \psi_{123}(x_1, x_2, x_3) \cdot \psi_{234}(x_2, x_3, x_4)$$

An approach to compute Φ' in time $\mathcal{O}(N^{\frac{3}{2}})$:

- The summation over X_4 is only relevant to the second bag factor
- First aggregate away X_4 over ψ_{234}, then join and finally aggregate away X_2

$$\Phi''(x_1, x_3) = \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \left(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \right) \right)$$

Steps 1.3 + 2 Compute Φ'' by Marginalising the Bound Variables

$$\Phi''(x_1, x_3) = \sum_{x_2} \Big(\psi_{123}(x_1, x_2, x_3) \cdot \Big(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \Big) \Big)$$

Steps 1.3 + 2 Compute Φ'' by Marginalising the Bound Variables

$$\Phi''(x_1, x_3) = \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \left(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \right) \right)$$

$$= \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \underbrace{\left(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \right)}_{\psi'_{23}(x_2, x_3)} \right)$$

Steps 1.3 + 2 Compute Φ'' by Marginalising the Bound Variables

$$\Phi''(x_1, x_3) = \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \left(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \right) \right)$$

$$= \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \underbrace{\left(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \right)}_{\psi'_{23}(x_2, x_3)} \right)$$

$$= \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \psi'_{23}(x_2, x_3) \right) \qquad O(N^{3/2})$$

Steps 1.3 + 2 Compute Φ'' by Marginalising the Bound Variables

$$\Phi''(x_1, x_3) = \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \left(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \right) \right)$$

$$= \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \underbrace{\left(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \right)}_{\psi'_{23}(x_2, x_3)} \right)$$

$$= \sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \psi'_{23}(x_2, x_3) \right) \qquad O(N^{3/2})$$

$$= \underbrace{\sum_{x_2} \left(\psi_{123}(x_1, x_2, x_3) \cdot \psi'_{23}(x_2, x_3) \right)}_{\psi'_{123}(x_1, x_3)}$$

Steps 1.3 + 2 Compute Φ'' by Marginalising the Bound Variables

$$\Phi''(x_1, x_3) = \sum_{x_2} \Big(\psi_{123}(x_1, x_2, x_3) \cdot \Big(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \Big) \Big)$$

$$= \sum_{x_2} \Big(\psi_{123}(x_1, x_2, x_3) \cdot \underbrace{\Big(\sum_{x_4} \psi_{234}(x_2, x_3, x_4) \Big)}_{\psi'_{23}(x_2, x_3)} \Big)$$

$$= \sum_{x_2} \Big(\psi_{123}(x_1, x_2, x_3) \cdot \psi'_{23}(x_2, x_3) \Big) \qquad O(N^{3/2})$$

$$= \underbrace{\sum_{x_2} \Big(\psi_{123}(x_1, x_2, x_3) \cdot \psi'_{23}(x_2, x_3) \Big)}_{\psi'_{123}(x_1, x_3)}$$

$$= \psi'_{13}(x_1, x_3) \qquad O(N^{3/2})$$

General Case

Query Rewriting: Indicator Projections

Example

Factor ψ_S

1	2	3	ψ_S
1	3	3	5
2	3	3	2
3	1	2	4

$T = \{2, 3, 4\}$

Indicator projection of ψ_S onto T

2	3	$\psi_{S/T}$
3	3	**1**
1	2	**1**

Query Rewriting: Indicator Projections

Example

Factor ψ_S

1	2	3	ψ_S
1	3	3	5
2	3	3	2
3	1	2	4

$T = \{2, 3, 4\}$

Indicator projection of ψ_S onto T

2	3	$\psi_{S/T}$
3	3	**1**
1	2	**1**

Let ψ_S be a factor and T a set with $S \cap T \neq \emptyset$

$$\psi_{S/T}(\mathbf{x}_{S\cap T}) = \begin{cases} \mathbf{1}, & \text{if } \exists \mathbf{x}_{S-T} \text{ s.t. } \psi_S(\mathbf{x}_{S-T}, \mathbf{x}_{S\cap T}) \neq \mathbf{0} \\ \mathbf{0}, & \text{otherwise} \end{cases}$$

is the indicator projection of ψ_S onto T

Which Copies of a Factor Become Indicator Projections?

Factor ψ_S can be used for computing several bags in the decomposition $\mathcal{T}$

- Consider each bag B_i in $\mathcal{T}$ with variables T such that $S \cap T \neq \emptyset$
- If B_i is the highest bag of $\mathcal{T}$ such that $S \subseteq T$, then include the actual factor ψ_S into the computation of the bag factor ψ_{B_i}
- Otherwise, include the indicator projection ψ_{S/B_i} into the computation of the bag factor ψ_{B_i}

The above strategy is just a convention, any other convention is OK as long as the factor ψ_S is only used once for computing the bags in $\mathcal{T}$

Query Rewriting: Marginalisation Orders Compatible with Decompositions

Consider a hypertree decomposition $\mathcal{T}$

- A bag B of $\mathcal{T}$ owns a variable i if B is the highest bag in $\mathcal{T}$ containing i

Query Rewriting: Marginalisation Orders Compatible with Decompositions

Consider a hypertree decomposition $\mathcal{T}$

- A bag B of $\mathcal{T}$ owns a variable i if B is the highest bag in $\mathcal{T}$ containing i
- Let $B_1, \ldots, B_m$ be an ordering of the bags of $\mathcal{T}$ that is compatible with the partial order given by $\mathcal{T}$
    - The ordering $B_1, \ldots, B_m$ can be obtained by a pre-order traversal of $\mathcal{T}$

Query Rewriting: Marginalisation Orders Compatible with Decompositions

Consider a hypertree decomposition $\mathcal{T}$

- A bag B of $\mathcal{T}$ owns a variable i if B is the highest bag in $\mathcal{T}$ containing i
- Let $B_1, \ldots, B_m$ be an ordering of the bags of $\mathcal{T}$ that is compatible with the partial order given by $\mathcal{T}$
  - The ordering $B_1, \ldots, B_m$ can be obtained by a pre-order traversal of $\mathcal{T}$
- For $i \in [m]$, let σ_i be the list of bound variables owned by B_i in any order

Query Rewriting: Marginalisation Orders Compatible with Decompositions

Consider a hypertree decomposition $\mathcal{T}$

- A bag B of $\mathcal{T}$ owns a variable i if B is the highest bag in $\mathcal{T}$ containing i
- Let $B_1, \ldots, B_m$ be an ordering of the bags of $\mathcal{T}$ that is compatible with the partial order given by $\mathcal{T}$
  - The ordering $B_1, \ldots, B_m$ can be obtained by a pre-order traversal of $\mathcal{T}$
- For $i \in [m]$, let σ_i be the list of bound variables owned by B_i in any order
- $\sigma = \sigma_1 \ldots \sigma_m$ is a marginalisation order compatible with $\mathcal{T}$

Query Rewriting: Eagerly Moving Sums Past Products

Input:

- FAQ-SS $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}} \cdots \bigoplus_{x_{n-1}} \bigoplus_{x_n} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$
- Marginalisation order $\sigma = (f+1, \ldots, n)$

Query Rewriting: Eagerly Moving Sums Past Products

Input:

- FAQ-SS $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}} \cdots \bigoplus_{x_{n-1}} \bigoplus_{x_n} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$
- Marginalisation order $\sigma = (f+1, \dots, n)$

We rewrite Φ by repeatedly moving the sums past products

- Move $\bigoplus_{x_n}$: Let $\partial(n) = \{S \in \mathcal{E} \mid n \in S\}$ and $U = \bigcup_{S \in \partial(n)} (S \setminus \{n\})$

$$\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}} \cdots \bigoplus_{x_{n-1}} \bigotimes_{S \in \mathcal{E} \setminus \partial(n)} \psi_S(\mathbf{x}_S) \otimes \underbrace{\bigoplus_{x_n} \bigotimes_{S \in \partial(n)} \psi_S(\mathbf{x}_S)}_{\psi_U}$$

$$= \bigoplus_{x_{f+1}} \cdots \bigoplus_{x_{n-1}} \bigotimes_{S \in \mathcal{E} \setminus \partial(n)} \psi_S(\mathbf{x}_S) \otimes \psi_U(\mathbf{x}_U)$$

- We now have a new factor $\psi_U(\mathbf{x}_U)$ and a new hypergraph for Φ
- We next move $\bigoplus_{x_{n-1}}$ and so on until $\bigoplus_{x_f}$

Query Evaluation: Computing the FAQ-SS

Consider an FAQ-SS Φ where sums are moved past products

- Use Leapfrog Triejoin to compute each of the new factors ψ_U

We are left with an α-acyclic FAQ $\Phi(\mathbf{x}_{[f]}) = \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$ where all bound variables are marginalised

Query Evaluation: Computing the FAQ-SS

Consider an FAQ-SS Φ where sums are moved past products

- Use Leapfrog Triejoin to compute each of the new factors ψ_U

We are left with an α-acyclic FAQ $\Phi(\mathbf{x}_{[f]}) = \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$ where all bound variables are marginalised

- Use Yannakakis' algorithm to compute the join of all factors ψ_S in Φ

Further FAQ-SS Solver Example

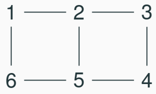
Example 1: The Grid (1/2)

$$\Phi() = \sum_{x_1, x_2, x_3, x_4, x_5, x_6} \psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{34}(x_3, x_4) \cdot \psi_{45}(x_4, x_5) \cdot \psi_{56}(x_5, x_6) \cdot \psi_{16}(x_1, x_6) \cdot \psi_{25}(x_2, x_5)$$

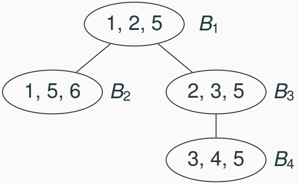
Example 1: The Grid (1/2)

$$\Phi() = \sum_{x_1, x_2, x_3, x_4, x_5, x_6} \psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_{34}(x_3, x_4) \cdot \psi_{45}(x_4, x_5) \cdot \psi_{56}(x_5, x_6) \cdot$$
$$\psi_{16}(x_1, x_6) \cdot \psi_{25}(x_2, x_5)$$

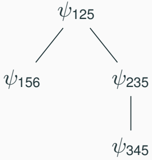
Hypergraph $\mathcal{H}$



Hypertree decomposition $\mathcal{T}$ for $\mathcal{H}$

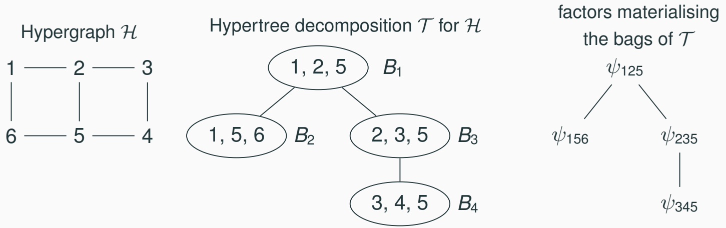


factors materialising the bags of $\mathcal{T}$



Example 1: The Grid (1/2)

$$\Phi() = \sum_{x_1,x_2,x_3,x_4,x_5,x_6} \psi_{12}(x_1,x_2) \cdot \psi_{23}(x_2,x_3) \cdot \psi_{34}(x_3,x_4) \cdot \psi_{45}(x_4,x_5) \cdot \psi_{56}(x_5,x_6) \cdot \psi_{16}(x_1,x_6) \cdot \psi_{25}(x_2,x_5)$$

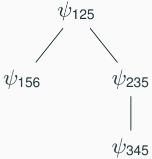


- Factor for B_1: $\psi_{125}(x_1, x_2, x_5) =$
  $\psi_{12}(x_1,x_2) \cdot \psi_{23/125}(x_2) \cdot \psi_{45/125}(x_5) \cdot \psi_{56/125}(x_5) \cdot \psi_{16/125}(x_1) \cdot \psi_{25}(x_2,x_5)$
- Factor for B_2: $\psi_{156}(x_1, x_5, x_6) =$
  $\psi_{12/156}(x_1) \cdot \psi_{45/156}(x_5) \cdot \psi_{56}(x_5,x_6) \cdot \psi_{16}(x_1,x_6) \cdot \psi_{25/156}(x_5)$
- Factor for B_3: $\psi_{235}(x_2, x_3, x_5) =$
  $\psi_{12/235}(x_2) \cdot \psi_{23}(x_2,x_3) \cdot \psi_{34/235}(x_3) \cdot \psi_{45/235}(x_5) \cdot \psi_{56/235}(x_5) \cdot \psi_{25/235}(x_2,x_5)$
- Factor for B_4: $\psi_{345}(x_3, x_4, x_5) =$
  $\psi_{23/345}(x_3) \cdot \psi_{34}(x_3,x_4) \cdot \psi_{45}(x_4,x_5) \cdot \psi_{56/345}(x_5) \cdot \psi_{25/345}(x_5)$

Example 1: The Grid (2/2)

$$\Phi'() = \sum_{x_1, x_2, x_5, x_6, x_3, x_4} \psi_{125}(x_1, x_2, x_5) \cdot \psi_{156}(x_1, x_5, x_6) \cdot \psi_{235}(x_2, x_3, x_5) \cdot \psi_{345}(x_3, x_4, x_5)$$

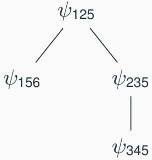
Marginalisation order: (1,2,5,6,3,4)



Example 1: The Grid (2/2)

$$\Phi'() = \sum_{x_1, x_2, x_5, x_6, x_3, x_4} \psi_{125}(x_1, x_2, x_5) \cdot \psi_{156}(x_1, x_5, x_6) \cdot \psi_{235}(x_2, x_3, x_5) \cdot \psi_{345}(x_3, x_4, x_5)$$

Marginalisation order: (1,2,5,6,3,4)



$$\Phi''() = \left(\sum_{x_1} \left(\sum_{x_2} \left(\sum_{x_5} \psi_{125}(x_1, x_2, x_5) \cdot \left(\sum_{x_6} \psi_{156}(x_1, x_5, x_6) \cdot \left(\sum_{x_3} \psi_{235}(x_2, x_3, x_5) \cdot \sum_{x_4} \psi_{345}(x_3, x_4, x_5) \right) \right) \right) \right) \right)$$

FAQs over Multiple Semirings

Differences to FAQ-SS

Consider an FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$

(1) The operators $\bigoplus^{(f+1)}, \ldots, \bigoplus^{(n)}$ can be from different semirings

- Application: e.g. Count SAT for quantified formulas (Exercise Sheet 1)

(2) We can have $\bigoplus^{(k)} = \bigotimes$ for some $k \in \{f+1, \ldots, n\}$

- Application: $\bigotimes$ simulates universal quantification (Exercise Sheet 1)

Differences to FAQ-SS

Consider an FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$

(1) The operators $\bigoplus^{(f+1)}, \ldots, \bigoplus^{(n)}$ can be from different semirings

- Application: e.g. Count SAT for quantified formulas (Exercise Sheet 1)

(2) We can have $\bigoplus^{(k)} = \bigotimes$ for some $k \in \{f+1, \ldots, n\}$

- Application: $\bigotimes$ simulates universal quantification (Exercise Sheet 1)

Marginalisation orders

- Property (1) restricts the number of possible marginalisation orders
- We cannot use arbitrary hypertree decompositions to derive marginalisation orders as in case of FAQ-SS

Differences to FAQ-SS

Consider an FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$

(1) The operators $\bigoplus^{(f+1)}, \ldots, \bigoplus^{(n)}$ can be from different semirings

- Application: e.g. Count SAT for quantified formulas (Exercise Sheet 1)

(2) We can have $\bigoplus^{(k)} = \bigotimes$ for some $k \in \{f+1, \ldots, n\}$

- Application: $\bigotimes$ simulates universal quantification (Exercise Sheet 1)

Marginalisation orders

- Property (1) restricts the number of possible marginalisation orders
- We cannot use arbitrary hypertree decompositions to derive marginalisation orders as in case of FAQ-SS

Product aggregates

- Marginalising X_k with $\bigoplus^{(k)} = \bigotimes$ can be easier than in case $\bigoplus^{(k)} \neq \bigotimes$

Marginalisation Orders

Swapping Aggregate Operators

Consider an FAQ

$$\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$$

- A marginalisation order $(i_1, \ldots, i_n)$ dictates that we first marginalise X_{i_n}, then $X_{i_{n-1}}$, and so on until X_{i_1}

Swapping Aggregate Operators

Consider an FAQ

$$\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$$

- A marginalisation order $(i_1, \ldots, i_n)$ dictates that we first marginalise X_{i_n}, then $X_{i_{n-1}}$, and so on until X_{i_1}
- A marginalisation order $(i_1, \ldots, i_n)$ is valid for Φ if

$$\Phi'(\mathbf{x}_{[f]}) = \bigoplus_{x_{i_1}}^{(i_1)} \cdots \bigoplus_{x_{i_n}}^{(i_n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$$

  is equivalent to Φ for all possible definitions of input factors $(\psi_S)_{S \in \mathcal{E}}$

Swapping Aggregate Operators

Consider an FAQ

$$\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S\in\mathcal{E}} \psi_S(\mathbf{x}_S)$$

- A marginalisation order $(i_1, \ldots, i_n)$ dictates that we first marginalise X_{i_n}, then $X_{i_{n-1}}$, and so on until X_{i_1}
- A marginalisation order $(i_1, \ldots, i_n)$ is valid for Φ if

$$\Phi'(\mathbf{x}_{[f]}) = \bigoplus_{x_{i_1}}^{(i_1)} \cdots \bigoplus_{x_{i_n}}^{(i_n)} \bigotimes_{S\in\mathcal{E}} \psi_S(\mathbf{x}_S)$$

  is equivalent to Φ for all possible definitions of input factors $(\psi_S)_{S\in\mathcal{E}}$

Not all marginalisation orders are valid for an FAQ

Example: Swapping Aggregate Operators not Always Correct

Consider the factor ψ_{12} and the two FAQs Φ_1 and Φ_2

X_1	X_2	ψ_{12}
1	2	2
1	3	4
2	2	3

$$\Phi_1() = \sum_{x_1} \max_{x_2} \psi_{12}(x_1, x_2) \qquad \Phi_2() = \max_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2)$$

Example: Swapping Aggregate Operators not Always Correct

Consider the factor ψ_{12} and the two FAQs Φ_1 and Φ_2

X_1	X_2	ψ_{12}
1	2	2
1	3	4
2	2	3

$$\Phi_1() = \sum_{x_1} \max_{x_2} \psi_{12}(x_1, x_2) \qquad \Phi_2() = \max_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2)$$

X_1	$\max_{x_2} \psi_{12}$
1	4
2	3

result of Φ_1

$()$	$\sum_{x_1} \max_{x_2} \psi_{12}$
$()$	7

Example: Swapping Aggregate Operators not Always Correct

Consider the factor ψ_{12} and the two FAQs Φ_1 and Φ_2

X_1	X_2	ψ_{12}
1	2	2
1	3	4
2	2	3

$$\Phi_1() = \sum_{x_1} \max_{x_2} \psi_{12}(x_1, x_2) \qquad \Phi_2() = \max_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2)$$

X_1	$\max_{x_2} \psi_{12}$
1	4
2	3

result of Φ_1

$()$	$\sum_{x_1} \max_{x_2} \psi_{12}$
$()$	7

X_2	$\sum_{x_1} \psi_{12}$
2	5
3	4

result of Φ_2

$()$	$\max_{x_2} \sum_{x_1} \psi_{12}$
$()$	5

Example: Swapping Aggregate Operators not Always Correct

Consider the factor ψ_{12} and the two FAQs Φ_1 and Φ_2

X_1	X_2	ψ_{12}
1	2	2
1	3	4
2	2	3

$$\Phi_1() = \sum_{x_1} \max_{x_2} \psi_{12}(x_1, x_2) \qquad \Phi_2() = \max_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2)$$

X_1	$\max_{x_2} \psi_{12}$
1	4
2	3

result of Φ_1

$()$	$\sum_{x_1} \max_{x_2} \psi_{12}$
$()$	7

X_2	$\sum_{x_1} \psi_{12}$
2	5
3	4

result of Φ_2

$()$	$\max_{x_2} \sum_{x_1} \psi_{12}$
$()$	5

Φ_1 and Φ_2 are not equivalent

Example: Marginalisation Orders Affect Evaluation Time

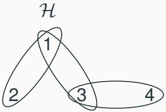
- Marginalisation orders affect the evaluation time of FAQs

Example: Marginalisation Orders Affect Evaluation Time

- Marginalisation orders affect the evaluation time of FAQs

- Consider the FAQ Φ with hypergraph $\mathcal{H}$

$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

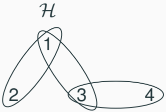


- The marginalisation orders $\sigma_1 = (3, 2, 1)$ and $\sigma_2 = (3, 1, 2)$ are valid for Φ

Example: Marginalisation Orders Affect Evaluation Time

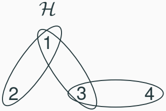
- Marginalisation orders affect the evaluation time of FAQs
- Consider the FAQ Φ with hypergraph $\mathcal{H}$

$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$



- The marginalisation orders $\sigma_1 = (3, 2, 1)$ and $\sigma_2 = (3, 1, 2)$ are valid for Φ
- We will see that σ_1 and σ_2 lead to different evaluation times

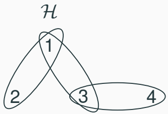
Example: A Bad Marginalisation Order



$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$

Marginalisation order: $\sigma_1 = (3, 2, 1)$

Example: A Bad Marginalisation Order

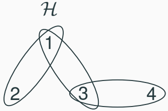


$\partial(1) = \{\{1, 2\}\, \{1, 3\}\}$ The hyperedges containing 1

$U = \{1, 2, 3\}$ Set of nodes occurring in the hyperedges containing 1

$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

Example: A Bad Marginalisation Order



$\partial(1) = \{\{1,2\}\,\{1,3\}\}$ The hyperedges containing 1

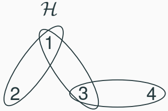
$U = \{1,2,3\}$ Set of nodes occurring in the hyperedges containing 1

$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_2} \psi_{34}(x_3, x_4) \cdot \underbrace{\left(\sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \right)}_{\psi'_{23}} \qquad \rho^*_{\mathcal{H}}(\{1,2,3\}) = 2$$

$\rho^*_{\mathcal{H}}(\{1,2,3\})$: ρ^* for the subgraph of $\mathcal{H}$ induced by the nodes $\{1,2,3\}$

Example: A Bad Marginalisation Order



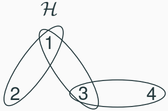
$\partial(1) = \{\{1,2\}\,\{1,3\}\}$

$U = \{1,2,3\}$

$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$ compute time

$= \max_{x_3} \sum_{x_2} \psi_{34}(x_3, x_4) \cdot \psi'_{23}(x_2, x_3)$ $\mathcal{O}(N^2)$

Example: A Bad Marginalisation Order



$\partial(1) = \{\{1,2\}\,\{1,3\}\}$ $\quad U = \{1,2,3\}$

$\partial(2) = \{\{2,3\}\}$ $\quad U = \{2,3\}$

$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

compute time

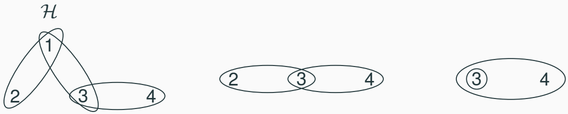
$$= \max_{x_3} \sum_{x_2} \psi_{34}(x_3, x_4) \cdot \psi'_{23}(x_2, x_3)$$

$\mathcal{O}(N^2)$

$$= \max_{x_3} \psi_{34}(x_3, x_4) \cdot \underbrace{\left(\sum_{x_2} \psi'_{23}(x_2, x_3)\right)}_{\psi'_3}$$

$\rho^*_{\mathcal{H}}(\{2,3\}) = 2$

Example: A Bad Marginalisation Order



$\partial(1) = \{\{1,2\}\,\{1,3\}\}$ $\quad$ $\partial(2) = \{\{2,3\}\}$

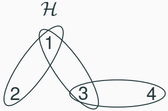
$U = \{1,2,3\}$ $\quad$ $U = \{2,3\}$

$$\Phi(x_4) = \max_{x_3} \textstyle\sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4) \qquad \text{compute time}$$

$$= \max_{x_3} \textstyle\sum_{x_2} \psi_{34}(x_3, x_4) \cdot \psi'_{23}(x_2, x_3) \qquad \mathcal{O}(N^2)$$

$$= \max_{x_3} \psi_{34}(x_3, x_4) \cdot \psi'_3(x_3) \qquad \mathcal{O}(N^2)$$

Example: A Bad Marginalisation Order



$\partial(1) = \{\{1,2\}\ \{1,3\}\}$ $\quad U = \{1,2,3\}$

$\partial(2) = \{\{2,3\}\}$ $\quad U = \{2,3\}$

$\partial(3) = \{\{3\},\{3,4\}\}$ $\quad U = \{3,4\}$

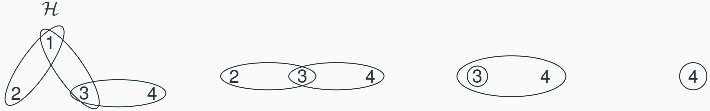
$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1,x_2) \cdot \psi_{13}(x_1,x_3) \cdot \psi_{34}(x_3,x_4)$$ compute time

$$= \max_{x_3} \sum_{x_2} \psi_{34}(x_3,x_4) \cdot \psi'_{23}(x_2,x_3)$$ $\mathcal{O}(N^2)$

$$= \max_{x_3} \psi_{34}(x_3,x_4) \cdot \psi'_3(x_3)$$ $\mathcal{O}(N^2)$

$$= \underbrace{\left(\max_{x_3} \psi_{34}(x_4,x_3) \cdot \psi'_3(x_3) \right)}_{\psi'_4}$$ $\rho^*_{\mathcal{H}}(\{3,4\}) = 1$

Example: A Bad Marginalisation Order

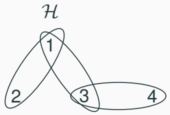


$\partial_4(1) = \{\{1,2\}\,\{1,3\}\}$ $\quad$ $\partial(2) = \{\{2,3\}\}$ $\quad$ $\partial(3) = \{\{3\},\{3,4\}\}$

$U_4 = \{1,2,3\}$ $\quad$ $U = \{2,3\}$ $\quad$ $U = \{3,4\}$

$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$	compute time
$= \max_{x_3} \sum_{x_2} \psi_{34}(x_3, x_4) \cdot \psi'_{23}(x_2, x_3)$	$\mathcal{O}(N^2)$
$= \max_{x_3} \psi_{34}(x_3, x_4) \cdot \psi'_3(x_3)$	$\mathcal{O}(N^2)$
$= \psi'_4(x_4)$	$\mathcal{O}(N)$
	overall $\mathcal{O}(N^2)$

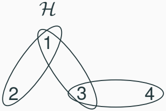
Example: A Better Marginalisation Order



$\Phi(x_4) = \max_{x_3} \sum_{x_1} \sum_{x_2} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$

Marginalisation order: $\sigma_1 = (3, 1, 2)$

Example: A Better Marginalisation Order



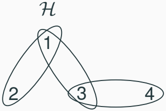
$\partial(2) = \{\{1, 2\}\}$

$U = \{1, 2\}$

$$\Phi(x_4) = \max_{x_3} \sum_{x_1} \sum_{x_2} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_1} \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4) \cdot \underbrace{\left(\sum_{x_2} \psi_{12}(x_1, x_2) \right)}_{\psi'_1} \qquad \rho^*_{\mathcal{H}}(\{1, 2\}) = 1$$

Example: A Better Marginalisation Order



$$\partial(2) = \{\{1, 2\}\}$$

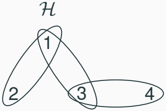
$$U = \{1, 2\}$$

compute time

$$\Phi(x_4) = \max_{x_3} \sum_{x_1} \sum_{x_2} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_1} \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4) \cdot \psi_1'(x_1) \qquad \mathcal{O}(N)$$

Example: A Better Marginalisation Order



$\partial(2) = \{\{1, 2\}\}$ $\quad$ $\partial(1) = \{\{1\}, \{1, 3\}\}$

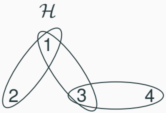
$U = \{1, 2\}$ $\quad$ $U = \{1, 3\}$

$$\Phi(x_4) = \max_{x_3} \sum_{x_1} \sum_{x_2} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4) \qquad \text{compute time}$$

$$= \max_{x_3} \sum_{x_1} \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4) \cdot \psi'_1(x_1) \qquad \mathcal{O}(N)$$

$$= \max_{x_3} \psi_{34}(x_3, x_4) \cdot \underbrace{\left(\sum_{x_1} \psi_{13}(x_1, x_3) \cdot \psi'_1(x_1) \right)}_{\psi'_3} \qquad \rho^*_{\mathcal{H}}(\{1, 3\}) = 1$$

Example: A Better Marginalisation Order



$\partial(2) = \{\{1,2\}\}$

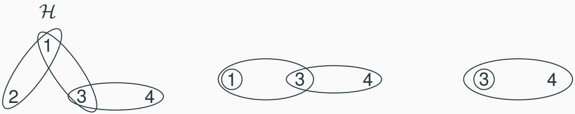
$U = \{1,2\}$

$\partial(1) = \{\{1\},\{1,3\}\}$

$U = \{1,3\}$

	compute time
$\Phi(x_4) = \max_{x_3} \sum_{x_1} \sum_{x_2} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$	
$= \max_{x_3} \sum_{x_1} \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4) \cdot \psi'_1(x_1)$	$\mathcal{O}(N)$
$= \max_{x_3} \psi_{34}(x_3, x_4) \cdot \psi'_3(x_3)$	$\mathcal{O}(N)$

Example: A Better Marginalisation Order



$\partial(2) = \{\{1,2\}\}$ $\quad$ $\partial(1) = \{\{1\},\{1,3\}\}$ $\quad$ $\partial(3) = \{\{3\},\{3,4\}\}$

$U = \{1,2\}$ $\quad$ $U = \{1,3\}$ $\quad$ $U = \{3,4\}$

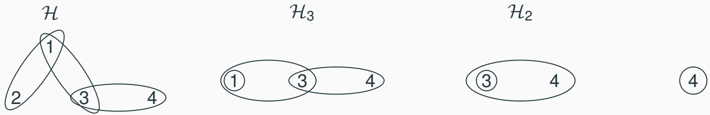
$$\Phi(x_4) = \max_{x_3} \sum_{x_1} \sum_{x_2} \psi_{12}(x_1,x_2)\cdot\psi_{13}(x_1,x_3)\cdot\psi_{34}(x_3,x_4)$$ compute time

$$= \max_{x_3} \sum_{x_1} \psi_{13}(x_1,x_3)\cdot\psi_{34}(x_3,x_4)\cdot\psi'_1(x_1) \qquad \mathcal{O}(N)$$

$$= \max_{x_3} \psi_{34}(x_3,x_4)\cdot\psi'_3(x_3) \qquad \mathcal{O}(N)$$

$$= \underbrace{\left(\max_{x_3} \psi_{34}(x_3,x_4)\cdot\psi'_3(x_3)\right)}_{\psi'_4} \qquad \rho^*_{\mathcal{H}}(\{3,4\}) = 1$$

Example: A Better Marginalisation Order



$\partial(2) = \{\{1,2\}\}$ $\quad$ $\partial(1) = \{\{1\},\{1,3\}\}$ $\quad$ $\partial(3) = \{\{3\},\{3,4\}\}$

$U = \{1,2\}$ $\quad$ $U = \{1,3\}$ $\quad$ $U = \{3,4\}$

	compute time
$\Phi(x_4) = \max_{x_3} \sum_{x_1} \sum_{x_2} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$	
$\max_{x_3} \sum_{x_1} \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4) \cdot \psi_1'(x_1)$	$\mathcal{O}(N)$
$\max_{x_3} \psi_{34}(x_3, x_4) \cdot \psi_3'(x_3)$	$\mathcal{O}(N)$
$= \psi_4'(x_4)$	$\mathcal{O}(N)$
	overall $\mathcal{O}(N)$

Finding the Best Marginalisation Order

- Different marginalisation orders can lead to different evaluation times
- To find the optimal marginalisation order we need to go over the entire space of valid marginalisation orders
- In the general case, the existing algorithms are not better than enumerating over all possible orders
- Focus of this lecture: What is the evaluation time for a given FAQ and marginalisation order?

Product Aggregates

Example: Moving Product Marginalisation Past Factor Products 1/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3) \Big)$$

Example: Moving Product Marginalisation Past Factor Products 1/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3)\Big)$$

Assume $\text{Dom}(X_2) = \{2, 3\}$

$$\Phi(1, 2) = \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2)\Big)$$

Example: Moving Product Marginalisation Past Factor Products 1/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3) \Big)$$

Assume $\text{Dom}(X_2) = \{2, 3\}$

$$\Phi(1, 2) = \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2) \Big)$$

$$= (2 \cdot 4 \cdot 3) \cdot (3 \cdot 5 \cdot 3)$$

Example: Moving Product Marginalisation Past Factor Products 1/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3) \Big)$$

Assume $\text{Dom}(X_2) = \{2, 3\}$

$$\begin{aligned} \Phi(1, 2) &= \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2) \Big) \\ &= (2 \cdot 4 \cdot 3) \cdot (3 \cdot 5 \cdot 3) \\ &= (2 \cdot 3) \cdot (4 \cdot 5) \cdot (3 \cdot 3) \end{aligned}$$

Example: Moving Product Marginalisation Past Factor Products 1/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3)\Big)$$

Assume $\text{Dom}(X_2) = \{2, 3\}$

$$\begin{aligned}
\Phi(1,2) &= \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2)\Big) \\
&= (2 \cdot 4 \cdot 3) \cdot (3 \cdot 5 \cdot 3) \\
&= (2 \cdot 3) \cdot (4 \cdot 5) \cdot (3 \cdot 3) \\
&= \prod_{x_2} \psi_{12}(1, x_2) \cdot \prod_{x_2} \psi_{23}(x_2, 2) \cdot \psi_3(2)^{|\text{Dom}(X_2)|}
\end{aligned}$$

Example: Moving Product Marginalisation Past Factor Products 2/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3)\Big)$$

Assume now $\text{Dom}(X_2) = \{1, 2, 3\}$

$$\Phi(1, 2) = \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2)\Big)$$

Example: Moving Product Marginalisation Past Factor Products 2/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3)\Big)$$

Assume now $\text{Dom}(X_2) = \{1, 2, 3\}$

$$\Phi(1, 2) = \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2)\Big)$$

$$= 0 \cdot (2 \cdot 4 \cdot 3) \cdot (3 \cdot 5 \cdot 3)$$

Example: Moving Product Marginalisation Past Factor Products 2/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3)\Big)$$

Assume now $\text{Dom}(X_2) = \{1, 2, 3\}$

$$\begin{aligned}\Phi(1, 2) &= \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2)\Big) \\ &= 0 \cdot (2 \cdot 4 \cdot 3) \cdot (3 \cdot 5 \cdot 3) \\ &= 0\end{aligned}$$

Example: Moving Product Marginalisation Past Factor Products 2/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3)\Big)$$

Assume now $\text{Dom}(X_2) = \{1, 2, 3\}$

$$\begin{aligned}\Phi(1, 2) &= \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2)\Big) \\ &= 0 \cdot (2 \cdot 4 \cdot 3) \cdot (3 \cdot 5 \cdot 3) \\ &= 0 \\ &= \prod_{x_2} \psi_{12}(1, x_2) \cdot \prod_{x_2} \psi_{23}(x_2, 2) \cdot \psi_3(2)^{|\text{Dom}(X_2)|}\end{aligned}$$

Example: Moving Product Marginalisation Past Factor Products 2/2

Consider input factors ψ_{12}, ψ_{23}, and ψ_3 and FAQ Φ

X_1	X_2	ψ_{12}
1	2	2
1	3	3

X_2	X_3	ψ_{23}
2	2	4
3	2	5

X_3	ψ_3
2	3
5	4

$$\Phi(x_1, x_3) = \prod_{x_2} \Big(\psi_{12}(x_1, x_2) \cdot \psi_{23}(x_2, x_3) \cdot \psi_3(x_3)\Big)$$

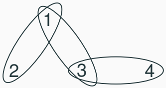
Assume now $\text{Dom}(X_2) = \{1, 2, 3\}$

$$\begin{aligned}
\Phi(1, 2) &= \prod_{x_2} \Big(\psi_{12}(1, x_2) \cdot \psi_{23}(x_2, 2) \cdot \psi_3(2)\Big) \\
&= 0 \cdot (2 \cdot 4 \cdot 3) \cdot (3 \cdot 5 \cdot 3) \\
&= 0 \\
&= \prod_{x_2} \psi_{12}(1, x_2) \cdot \prod_{x_2} \psi_{23}(x_2, 2) \cdot \psi_3(2)^{|\text{Dom}(X_2)|}
\end{aligned}$$

- We will use this rewriting strategy to deal with product aggregates

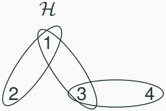
Product Aggregates

$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \sum_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$



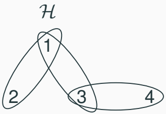
- Consider again the marginalisation order $(3, 2, 1)$ for Φ_4
- We have seen: The marginalisation of X_1 requires $\mathcal{O}(N^2)$ compute time
- Assume now: $\sum_{x_1}$ is replaced by $\prod_{x_1}$
- We will see: In this case, X_1 can be marginalised in $\mathcal{O}(N)$ time (with an additional logarithmic factor)

Example: Product Aggregates



$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

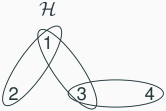
Example: Product Aggregates



$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}}_{\psi'_{34}} \cdot \Big(\prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3)\Big)$$

Example: Product Aggregates

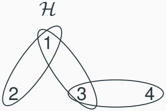


$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}}_{\psi'_{34}} \cdot \Big(\prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \Big)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)^{|\mathrm{Dom}(X_1)|}\Big)}_{\psi'_{34}} \cdot \Big(\underbrace{\prod_{x_1} \psi_{12}(x_1, x_2)}_{\psi'_2} \cdot \underbrace{\prod_{x_1} \psi_{13}(x_1, x_3)}_{\psi'_3} \Big)$$

Example: Product Aggregates



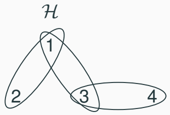
$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}}_{\psi'_{34}} \cdot \Big(\prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3)\Big)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)^{|\mathrm{Dom}(X_1)|}\Big)}_{\psi'_{34}} \cdot \Big(\underbrace{\prod_{x_1} \psi_{12}(x_1, x_2)}_{\psi'_2} \cdot \underbrace{\prod_{x_1} \psi_{13}(x_1, x_3)}_{\psi'_3} \Big)$$

- Compute time for ψ'_2 (same for ψ'_3): $\mathcal{O}(N)$ since we need to iterate over a single factor of size N

Example: Product Aggregates



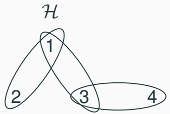
$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}}_{\psi'_{34}} \cdot \Big(\prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3)\Big)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)^{|\mathrm{Dom}(X_1)|}\Big)}_{\psi'_{34}} \cdot \Big(\underbrace{\prod_{x_1} \psi_{12}(x_1, x_2)}_{\psi'_2} \cdot \underbrace{\prod_{x_1} \psi_{13}(x_1, x_3)}_{\psi'_3} \Big)$$

- Compute time for ψ'_2 (same for ψ'_3): $\mathcal{O}(N)$ since we need to iterate over a single factor of size N
- $\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}$ is $\psi_{34}(x_3, x_4)$ to the power of $|\mathrm{Dom}(X_1)|$

Example: Product Aggregates



$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}}_{\psi'_{34}} \cdot \Big(\prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \Big)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)^{|\mathrm{Dom}(X_1)|}\Big)}_{\psi'_{34}} \cdot \Big(\underbrace{\prod_{x_1} \psi_{12}(x_1, x_2)}_{\psi'_2} \cdot \underbrace{\prod_{x_1} \psi_{13}(x_1, x_3)}_{\psi'_3} \Big)$$

- Compute time for ψ'_2 (same for ψ'_3): $\mathcal{O}(N)$ since we need to iterate over a single factor of size N
- $\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}$ is $\psi_{34}(x_3, x_4)$ to the power of $|\mathrm{Dom}(X_1)|$
- Next we look closer at the computation of $\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}$

Powering Factors

Consider a factor ψ_S

Powering Factors

Consider a factor ψ_S

- We need to compute a factor ψ'_S with $\psi'_S(\mathbf{x}_S) = \psi_S(\mathbf{x}_S)^n$ for each $\mathbf{x}_S$

Powering Factors

Consider a factor ψ_S

- We need to compute a factor ψ'_S with $\psi'_S(\mathbf{x}_S) = \psi_S(\mathbf{x}_S)^n$ for each $\mathbf{x}_S$
- ψ'_S has the same number of tuples as ψ_S

Powering Factors

Consider a factor ψ_S

- We need to compute a factor ψ'_S with $\psi'_S(\mathbf{x}_S) = \psi_S(\mathbf{x}_S)^n$ for each $\mathbf{x}_S$
- ψ'_S has the same number of tuples as ψ_S
- ψ'_S can be computed in $\mathcal{O}(|\psi_S| \cdot \log_2 n)$ time by repeated squaring

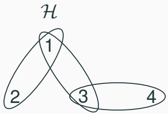
Powering Factors

Consider a factor ψ_S

- We need to compute a factor ψ'_S with $\psi'_S(\mathbf{x}_S) = \psi_S(\mathbf{x}_S)^n$ for each $\mathbf{x}_S$
- ψ'_S has the same number of tuples as ψ_S
- ψ'_S can be computed in $\mathcal{O}(|\psi_S| \cdot \log_2 n)$ time by repeated squaring
- If the range of ψ_S is idempotent, i.e. contains only 0 and 1, then $\psi'_S = \psi_S$

  $\Longrightarrow$ Compute time is $\mathcal{O}(1)$

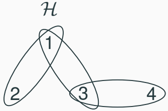
Example: Product Aggregates



$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$

Marginalisation order: $\sigma = (3, 2, 1)$

Example: Product Aggregates

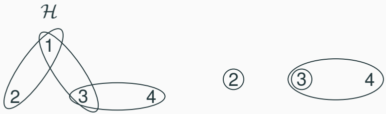


$\partial(1) = \{\{1, 2\}\,\{1, 3\}\}$

$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

$$= \max_{x_3} \sum_{x_2} \underbrace{\Big(\psi_{34}(x_3, x_4)\Big)^{|\mathrm{Dom}(X_1)|}}_{\psi'_{34}} \cdot \underbrace{\Big(\prod_{x_1} \psi_{12}(x_1, x_2)\Big)}_{\psi'_2} \cdot \underbrace{\Big(\prod_{x_1} \psi_{13}(x_1, x_3)\Big)}_{\psi'_3}$$

Example: Product Aggregates



$\partial(1) = \{\{1,2\}\,\{1,3\}\}$

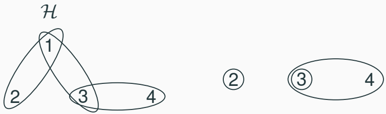
$$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$$

compute time

$$= \max_{x_3} \sum_{x_2} \psi'_{34}(x_3, x_4) \cdot \psi'_2(x_2) \cdot \psi'_3(x_3)$$

$\mathcal{O}(N)$

Example: Product Aggregates



$\partial(1) = \{\{1,2\}\,\{1,3\}\}$ $\qquad$ $\partial(2) = \{\{2\}\}$

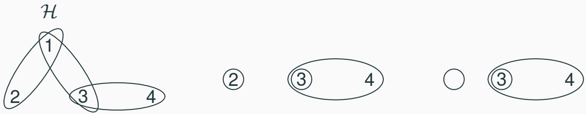
$U = \{2\}$

$$\Phi(x_4) = \max_{x_3} \textstyle\sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4) \qquad \text{compute time}$$

$$= \max_{x_3} \textstyle\sum_{x_2} \psi'_{34}(x_3, x_4) \cdot \psi'_2(x_2) \cdot \psi'_3(x_3) \qquad \mathcal{O}(N)$$

$$= \max_{x_3} \psi'_{34}(x_3, x_4) \cdot \psi'_3(x_3) \cdot \underbrace{\left(\sum_{x_2} \psi'_2(x_2) \right)}_{\psi'} \qquad \rho^*_{\mathcal{H}}(\{2\}) = 1$$

Example: Product Aggregates

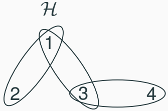


$\partial(1) = \{\{1,2\}\,\{1,3\}\}$ $\quad\quad$ $\partial(2) = \{\{2\}\}$

$U = \{2\}$

$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1, x_2) \cdot \psi_{13}(x_1, x_3) \cdot \psi_{34}(x_3, x_4)$	compute time
$= \max_{x_3} \sum_{x_2} \psi'_{34}(x_3, x_4) \cdot \psi'_2(x_2) \cdot \psi'_3(x_3)$	$\mathcal{O}(N)$
$= \max_{x_3} \psi'_{34}(x_3, x_4) \cdot \psi'_3(x_3) \cdot \psi'()$	$\mathcal{O}(N)$

Example: Product Aggregates



$\partial(1) = \{\{1,2\}\,\{1,3\}\}$

$\partial(2) = \{\{2\}\}$

$U = \{2\}$

$\partial(3) = \{\{3\},\{3,4\}\}$

$U = \{3,4\}$

	compute time
$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1,x_2) \cdot \psi_{13}(x_1,x_3) \cdot \psi_{34}(x_3,x_4)$	
$= \max_{x_3} \sum_{x_2} \psi'_{34}(x_3,x_4) \cdot \psi'_2(x_2) \cdot \psi'_3(x_3)$	$\mathcal{O}(N)$
$= \max_{x_3} \psi'_{34}(x_3,x_4) \cdot \psi'_3(x_3) \cdot \psi'()$	$\mathcal{O}(N)$
$= \psi'() \cdot \underbrace{\left(\max_{x_3} \psi'_{34}(x_3,x_4) \cdot \psi'_3(x_3)\right)}_{\psi'_4}$	$\rho^*_{\mathcal{H}}(\{3,4\}) = 1$

Example: Product Aggregates

$\partial_4(1) = \{\{1,2\}\,\{1,3\}\}$

$\partial(2) = \{\{2\}\}$

$U = \{2\}$

$\partial(3) = \{\{3\},\{3,4\}\}$

$U = \{3,4\}$

$\Phi(x_4) = \max_{x_3} \sum_{x_2} \prod_{x_1} \psi_{12}(x_1,x_2) \cdot \psi_{13}(x_1,x_3) \cdot \psi_{34}(x_3,x_4)$	compute time
$= \max_{x_3} \sum_{x_2} \psi'_{34}(x_3,x_4) \cdot \psi'_2(x_2) \cdot \psi'_3(x_3)$	$\mathcal{O}(N)$
$= \max_{x_3} \psi'_{34}(x_3,x_4) \cdot \psi'_3(x_3) \cdot \psi'()$	$\mathcal{O}(N)$
$= \psi'() \cdot \psi'_4(x_4)$	$\mathcal{O}(N)$

overall O(N)

InsideOut Algorithm

Main Steps of the InsideOut Algorithm

Input: FAQ Φ and marginalisation order σ for bound variables

Step 1: Marginalisation

- Marginalise the bound variables following σ
  $\Longrightarrow$ results in an FAQ with hypergraph $\mathcal{H}'$

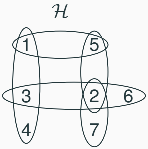
Step 2: Evaluation

- Construct a hypertree decomposition $\mathcal{T}$ for $\mathcal{H}'$
- Turn Φ into an α-acyclic FAQ Φ' by materialising the bags of $\mathcal{T}$
- Run Yannakakis' algorithm on Φ'

We next explain the above steps using an example

InsideOut Algorithm Example

Example: InsideOut Algorithm 1/3

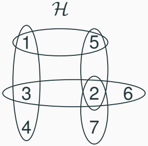


$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27}$

(for brevity, input variables of factors are skipped)

Marginalisation order: $\sigma = (3, 4, 5, 6)$

Example: InsideOut Algorithm 1/3



$\partial(6) = \{\{2, 3, 6\}\}$

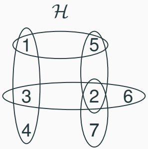
$U = \{2, 3, 6\}$

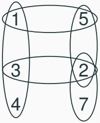
$$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27}$$

$$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \underbrace{\left(\max_{x_6} \underbrace{\psi_{25/236} \cdot \psi_{134/236} \cdot \psi_{27/236}}_{\text{indicator projections}} \cdot \psi_{236} \right)}_{\psi'_{23}}$$

$$\rho^*_{\mathcal{H}}(\{2, 3, 6\}) = 1$$

Example: InsideOut Algorithm 1/3





$$\partial(6) = \{\{2, 3, 6\}\}$$

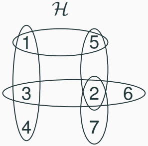
$$U = \{2, 3, 6\}$$

compute time

$$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27}$$

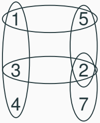
$$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \qquad \mathcal{O}(N)$$

Example: InsideOut Algorithm 1/3



$\partial(6) = \{\{2,3,6\}\}$

$U = \{2,3,6\}$



$\partial(5) = \{\{1,5\},\{2,5\}\}$

$U = \{1,2,5\}$

$$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27}$$

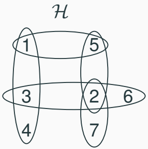
compute time

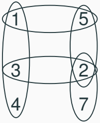
$$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \qquad \mathcal{O}(N)$$

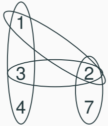
$$= \prod_{x_3} \sum_{x_4} \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \cdot \underbrace{\left(\max_{x_5} \psi_{134/125} \cdot \psi'_{23/125} \cdot \psi'_{27/125} \cdot \psi_{15} \cdot \psi_{25} \right)}_{\psi'_{12}}$$

$$\rho^*_{\mathcal{H}}(\{1,2,5\}) = 2$$

Example: InsideOut Algorithm 1/3







$\partial(6) = \{\{2,3,6\}\}$ $\quad$ $\partial(5) = \{\{1,5\},\{2,5\}\}$

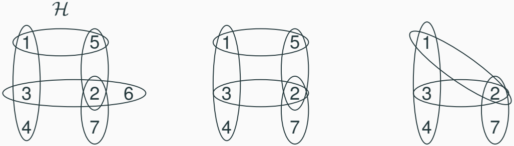
$U = \{2,3,6\}$ $\quad$ $U = \{1,2,5\}$

$$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27} \qquad \text{compute time}$$

$$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \qquad \mathcal{O}(N)$$

$$= \prod_{x_3} \sum_{x_4} \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12} \qquad \mathcal{O}(N^2)$$

Example: InsideOut Algorithm 1/3



$\partial(6) = \{\{2,3,6\}\}$ $\quad$ $\partial(5) = \{\{1,5\},\{2,5\}\}$ $\quad$ $\partial(4) = \{\{1,3,4\}\}$

$U = \{2,3,6\}$ $\quad$ $U = \{1,2,5\}$ $\quad$ $U = \{1,3,4\}$

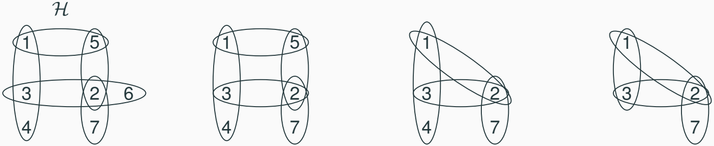
$$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27} \qquad \text{compute time}$$

$$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \qquad \mathcal{O}(N)$$

$$= \prod_{x_3} \sum_{x_4} \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12} \qquad \mathcal{O}(N^2)$$

$$= \prod_{x_3} \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12} \cdot \underbrace{\left(\sum_{x_4} \psi'_{23/134} \cdot \psi'_{12/134} \cdot \psi_{134} \right)}_{\psi'_{13}} \qquad \rho^*_{\mathcal{H}}(\{1,3,4\}) = 1$$

Example: InsideOut Algorithm 1/3



$\partial(6) = \{\{2,3,6\}\}$ $\quad$ $\partial(5) = \{\{1,5\},\{2,5\}\}$ $\quad$ $\partial(4) = \{\{1,3,4\}\}$

$U = \{2,3,6\}$ $\quad$ $U = \{1,2,5\}$ $\quad$ $U = \{1,3,4\}$

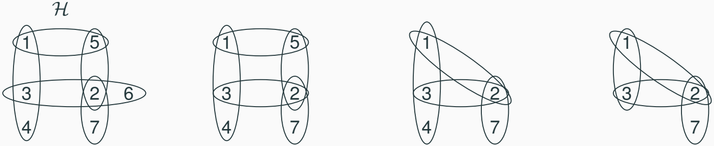
$$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27}$$ compute time

$$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \qquad \mathcal{O}(N)$$

$$= \prod_{x_3} \sum_{x_4} \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12} \qquad \mathcal{O}(N^2)$$

$$= \prod_{x_3} \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12} \cdot \psi'_{13} \qquad \mathcal{O}(N)$$

Example: InsideOut Algorithm 1/3



$\partial(6) = \{\{2,3,6\}\}$ $\quad$ $\partial(5) = \{\{1,5\},\{2,5\}\}$ $\quad$ $\partial(4) = \{\{1,3,4\}\}$ $\quad$ $\partial(3) = \{\{1,3\},\{2,3\}\}$

$U = \{2,3,6\}$ $\quad$ $U = \{1,2,5\}$ $\quad$ $U = \{1,3,4\}$

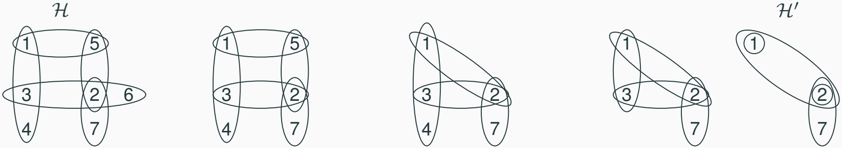
	compute time
$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27}$	
$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \psi'_{23}$	$\mathcal{O}(N)$
$= \prod_{x_3} \sum_{x_4} \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12}$	$\mathcal{O}(N^2)$
$= \prod_{x_3} \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12} \cdot \psi'_{13}$	$\mathcal{O}(N)$

$$= \underbrace{\left(\psi_{27}^{|\mathrm{Dom}(X_3)|}\right)}_{\psi'_{27}} \cdot \underbrace{\left(\psi'^{|\mathrm{Dom}(X_3)|}_{12}\right)}_{\psi''_{12}} \cdot \underbrace{\left(\prod_{x_3} \psi'_{23}\right)}_{\psi'_2} \cdot \underbrace{\left(\prod_{x_3} \psi'_{13}\right)}_{\psi'_1}$$

$\rho^*_{\mathcal{H}}(\{2,7\}) = \rho^*_{\mathcal{H}}(\{2,3\}) = \rho^*_{\mathcal{H}}(\{1,3\}) = 1$
sizes of ψ_{27}, ψ'_{23}, and ψ'_{13}: $\mathcal{O}(N)$
compute times for ψ'_{27}, ψ'_2, and ψ'_1: $\mathcal{O}(N)$

$\rho^*_{\mathcal{H}}(\{1,2\}) = 2$
size of ψ'_{12}: $\mathcal{O}(N^2)$
compute time for ψ''_{12}: $\mathcal{O}(N^2)$

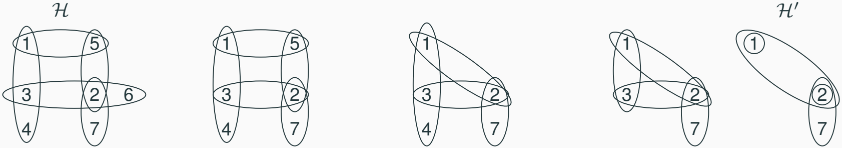
Example: InsideOut Algorithm 1/3



$\partial(6) = \{\{2,3,6\}\}$ $\quad$ $\partial(5) = \{\{1,5\},\{2,5\}\}$ $\quad$ $\partial(4) = \{\{1,3,4\}\}$ $\quad$ $\partial(3) = \{\{1,3\},\{2,3\}\}$

$U = \{2,3,6\}$ $\quad$ $U = \{1,2,5\}$ $\quad$ $U = \{1,3,4\}$

	compute time
$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27}$	
$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \psi'_{23}$	$\mathcal{O}(N)$
$= \prod_{x_3} \sum_{x_4} \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12}$	$\mathcal{O}(N^2)$
$= \prod_{x_3} \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12} \cdot \psi'_{13}$	$\mathcal{O}(N)$
$= \psi'_{27} \cdot \psi''_{12} \cdot \psi'_{2} \cdot \psi'_{1}$	$\mathcal{O}(N^2)$



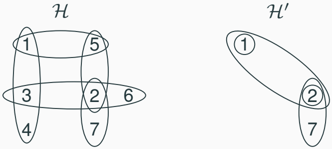
$\partial(6) = \{\{2,3,6\}\}$ $\quad$ $\partial(5) = \{\{1,5\},\{2,5\}\}$ $\quad$ $\partial(4) = \{\{1,3,4\}\}$ $\quad$ $\partial(3) = \{\{1,3\},\{2,3\}\}$

$U = \{2,3,6\}$ $\quad$ $U = \{1,2,5\}$ $\quad$ $U = \{1,3,4\}$

	compute time
$\Phi(x_1, x_2, x_7) = \prod_{x_3} \sum_{x_4} \max_{x_5} \max_{x_6} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{236} \cdot \psi_{27}$	
$= \prod_{x_3} \sum_{x_4} \max_{x_5} \psi_{15} \cdot \psi_{25} \cdot \psi_{134} \cdot \psi_{27} \cdot \psi'_{23}$	$\mathcal{O}(N)$
$= \prod_{x_3} \sum_{x_4} \psi_{134} \cdot \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12}$	$\mathcal{O}(N^2)$
$= \prod_{x_3} \psi_{27} \cdot \psi'_{23} \cdot \psi'_{12} \cdot \psi'_{13}$	$\mathcal{O}(N)$
$= \psi'_{27} \cdot \psi''_{12} \cdot \psi'_{2} \cdot \psi'_{1}$	$\mathcal{O}(N^2)$

- We are left with an FAQ with hypergraph $\mathcal{H}'$ and all variables free

Example: InsideOut Algorithm 2/3

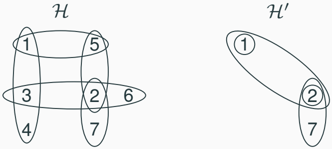


$$\Phi(x_1, x_2, x_7) = \psi'_{27}(x_2, x_7) \cdot \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi'_1(x_1)$$

After having marginalised all bound variables:

- Construct a hypertree decomposition $\mathcal{T}$ for $\mathcal{H}'$
- Materialise the bags of $\mathcal{T}$ using Leapfrog Triejoin (as in case of FAQ-SS)
- Compute the result of Φ using Yannakakis' algorithm

Example: InsideOut Algorithm 2/3



$$\Phi(x_1, x_2, x_7) = \psi'_{27}(x_2, x_7) \cdot \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi'_1(x_1)$$

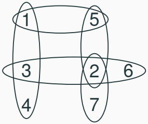
After having marginalised all bound variables:

- Construct a hypertree decomposition $\mathcal{T}$ for $\mathcal{H}'$
- Materialise the bags of $\mathcal{T}$ using Leapfrog Triejoin (as in case of FAQ-SS)
- Compute the result of Φ using Yannakakis' algorithm

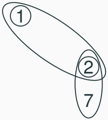
Next, we go through the above three steps.

Example: InsideOut Algorithm 3/3

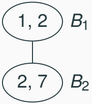
$\mathcal{H}$



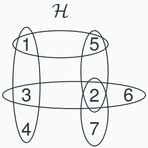
$\mathcal{H}'$

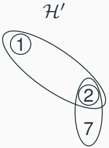


Hypertree decomposition $\mathcal{T}$ for $\mathcal{H}'$

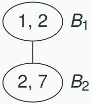


$$\Phi(x_1, x_2, x_7) = \psi'_{27}(x_2, x_7) \cdot \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi'_1(x_1)$$





Hypertree decomposition $\mathcal{T}$ for $\mathcal{H}'$

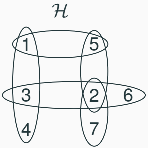


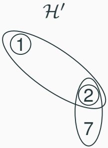
$$\Phi(x_1, x_2, x_7) = \psi'_{27}(x_2, x_7) \cdot \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi'_1(x_1)$$

Materialisation of the bags of $\mathcal{T}$

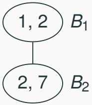
- $\psi_{B_1}(x_1, x_2) = \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi''_1(x_1) \cdot \psi'_{27/12}(x_2)$
  $\rho^*_{\mathcal{H}}(\{1, 2\}) = 2 \Rightarrow \mathcal{O}(N^2)$ compute time

Example: InsideOut Algorithm 3/3





Hypertree decomposition $\mathcal{T}$ for $\mathcal{H}'$



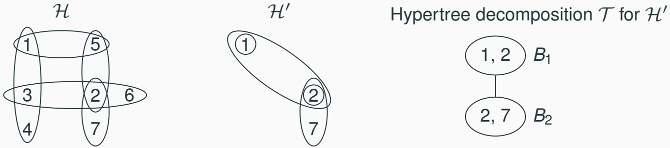
$$\Phi(x_1, x_2, x_7) = \psi'_{27}(x_2, x_7) \cdot \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi'_1(x_1)$$

Materialisation of the bags of $\mathcal{T}$

- $\psi_{B_1}(x_1, x_2) = \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi''_1(x_1) \cdot \psi'_{27/12}(x_2)$
  $\rho^*_{\mathcal{H}}(\{1, 2\}) = 2 \Rightarrow \mathcal{O}(N^2)$ compute time

- $\psi_{B_2}(x_2, x_7) = \psi'_{27} \cdot \psi''_{12/27}(x_2) \cdot \psi'_{2/27}(x_2)$
  $\rho^*_{\mathcal{H}}(\{2, 7\}) = 1 \Rightarrow \mathcal{O}(N)$ compute time

Example: InsideOut Algorithm 3/3



$$\Phi(x_1, x_2, x_7) = \psi'_{27}(x_2, x_7) \cdot \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi'_1(x_1)$$

Materialisation of the bags of $\mathcal{T}$

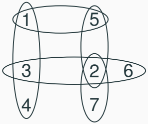
- $\psi_{B_1}(x_1, x_2) = \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi''_1(x_1) \cdot \psi'_{27/12}(x_2)$
  $\rho^*_{\mathcal{H}}(\{1, 2\}) = 2 \Rightarrow \mathcal{O}(N^2)$ compute time

- $\psi_{B_2}(x_2, x_7) = \psi'_{27} \cdot \psi''_{12/27}(x_2) \cdot \psi'_{2/27}(x_2)$
  $\rho^*_{\mathcal{H}}(\{2, 7\}) = 1 \Rightarrow \mathcal{O}(N)$ compute time

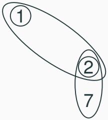
$$\Phi'(x_1, x_2, x_7) = \psi_{B_1}(x_1, x_2) \cdot \psi_{B_2}(x_2, x_7)$$

Example: InsideOut Algorithm 3/3

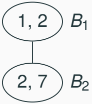
$\mathcal{H}$



$\mathcal{H}'$



Hypertree decomposition $\mathcal{T}$ for $\mathcal{H}'$



$$\Phi(x_1, x_2, x_7) = \psi'_{27}(x_2, x_7) \cdot \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi'_1(x_1)$$

Materialisation of the bags of $\mathcal{T}$

- $\psi_{B_1}(x_1, x_2) = \psi''_{12}(x_1, x_2) \cdot \psi'_2(x_2) \cdot \psi''_1(x_1) \cdot \psi'_{27/12}(x_2)$
  $\rho^*_{\mathcal{H}}(\{1,2\}) = 2 \Rightarrow \mathcal{O}(N^2)$ compute time

- $\psi_{B_2}(x_2, x_7) = \psi'_{27} \cdot \psi''_{12/27}(x_2) \cdot \psi'_{2/27}(x_2)$
  $\rho^*_{\mathcal{H}}(\{2,7\}) = 1 \Rightarrow \mathcal{O}(N)$ compute time

$$\Phi'(x_1, x_2, x_7) = \psi_{B_1}(x_1, x_2) \cdot \psi_{B_2}(x_2, x_7)$$

- Φ' is equivalent to Φ
- Compute the result of Φ' using Yannakakis' algorithm

InsideOut Algorithm in General

InsideOut Algorithm

Input:

- FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$
- Marginalisation order $\sigma = (f+1, \ldots, n)$

InsideOut Algorithm

Input:

- FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$
- Marginalisation order $\sigma = (f+1, \ldots, n)$

$\Phi_n = \Phi$
$\mathcal{H}_n = \mathcal{H}$

InsideOut Algorithm

Input:

- FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus^{(f+1)}_{x_{f+1}} \cdots \bigoplus^{(n)}_{x_n} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$
- Marginalisation order $\sigma = (f+1, \ldots, n)$

$\Phi_n = \Phi$
$\mathcal{H}_n = \mathcal{H}$

for $k = n$ down to $f+1$ do

if $\bigoplus^{(k)} \neq \bigotimes$

$(\Phi_{k-1}, \mathcal{H}_{k-1})$ = SemiringMarginalisation$(\Phi_k, \mathcal{H}_k)$

InsideOut Algorithm

Input:

- FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$
- Marginalisation order $\sigma = (f+1, \dots, n)$

$\Phi_n = \Phi$
$\mathcal{H}_n = \mathcal{H}$

for $k = n$ down to $f + 1$ do

if $\bigoplus^{(k)} \neq \bigotimes$

$(\Phi_{k-1}, \mathcal{H}_{k-1})$ = SemiringMarginalisation$(\Phi_k, \mathcal{H}_k)$

else

$(\Phi_{k-1}, \mathcal{H}_{k-1})$ = ProductMarginalisation$(\Phi_k, \mathcal{H}_k)$

InsideOut Algorithm

Input:

- FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$
- Marginalisation order $\sigma = (f+1, \ldots, n)$

$\Phi_n = \Phi$
$\mathcal{H}_n = \mathcal{H}$

for $k = n$ down to $f+1$ do

if $\bigoplus^{(k)} \neq \bigotimes$

$(\Phi_{k-1}, \mathcal{H}_{k-1})$ = SemiringMarginalisation$(\Phi_k, \mathcal{H}_k)$

else

$(\Phi_{k-1}, \mathcal{H}_{k-1})$ = ProductMarginalisation$(\Phi_k, \mathcal{H}_k)$

Construct for $\mathcal{H}_f$ a hypertree decomposition $\mathcal{T}$ with bags $(B_i)_{i \in m}$

Compute the bags $(\psi_{B_i})_{i \in [m]}$ of $\mathcal{T}$ using Leapfrog Triejoin

Run Yannakakis' algorithm on $\Phi'(\mathbf{x}_{[f]}) = \bigotimes_{i \in [m]} \psi_{B_i}(\mathbf{x}_{B_i})$

Procedure for Semiring Marginalisation

Procedure: SemiringMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

Procedure for Semiring Marginalisation

Procedure: SemiringMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

$\partial_k(k) = \{S \mid S \in \mathcal{E}_k \text{ with } k \in S\}$

$U_k = \bigcup_{S \in \partial_k(k)} S$

Procedure for Semiring Marginalisation

Procedure: SemiringMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

$\partial_k(k) = \{S \mid S \in \mathcal{E}_k \text{ with } k \in S\}$

$U_k = \bigcup_{S \in \partial_k(k)} S$

$$\psi_{U_k \setminus \{k\}}(\mathbf{x}_{U_k \setminus \{k\}}) = \bigoplus_{x_k} \left(\bigotimes_{S \in \partial_k(k)} \psi_S(\mathbf{x}_S) \otimes \bigotimes_{\substack{S \in \mathcal{E}_k \setminus \partial_k(k) \\ S \cap U_k \neq \emptyset}} \underbrace{\psi_{S/U_k}(\mathbf{x}_{S \cap U_k})}_{\text{indicator projection}} \right)$$

marginalisation

Procedure for Semiring Marginalisation

Procedure: SemiringMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

$\partial_k(k) = \{S \mid S \in \mathcal{E}_k \text{ with } k \in S\}$

$U_k = \bigcup_{S \in \partial_k(k)} S$

marginalisation

$$\psi_{U_k \setminus \{k\}}(\mathbf{x}_{U_k \setminus \{k\}}) = \bigoplus_{x_k} \left(\bigotimes_{S \in \partial_k(k)} \psi_S(\mathbf{x}_S) \otimes \bigotimes_{\substack{S \in \mathcal{E}_k \setminus \partial_k(k) \\ S \cap U_k \neq \emptyset}} \underbrace{\psi_{S/U_k}(\mathbf{x}_{S \cap U_k})}_{\text{indicator projection}} \right)$$

$\mathcal{V}_{k-1} = [k-1]$

$\mathcal{E}_{k-1} = (\mathcal{E}_k \setminus \partial_k(k)) \cup (U_k \setminus \{k\})$ (skip hyperedges $\partial_k(k)$, add hyperedge $U_k \setminus \{k\}$)

$\Phi_{k-1}(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_{k-1}}^{(k-1)} \bigotimes_{S \in \mathcal{E}_{k-1}} \psi_S(\mathbf{x}_S)$ (uses the new factor $\psi_{U_k \setminus \{k\}}$)

Procedure for Semiring Marginalisation

Procedure: SemiringMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

$\partial_k(k) = \{S \mid S \in \mathcal{E}_k \text{ with } k \in S\}$

$U_k = \bigcup_{S \in \partial_k(k)} S$

$$\psi_{U_k \setminus \{k\}}(\mathbf{x}_{U_k \setminus \{k\}}) = \bigoplus_{x_k} \Big(\bigotimes_{S \in \partial_k(k)} \psi_S(\mathbf{x}_S) \otimes \bigotimes_{\substack{S \in \mathcal{E}_k \setminus \partial_k(k) \\ S \cap U_k \neq \emptyset}} \underbrace{\psi_{S/U_k}(\mathbf{x}_{S \cap U_k})}_{\text{indicator projection}} \Big)$$

marginalisation

$\mathcal{V}_{k-1} = [k-1]$

$\mathcal{E}_{k-1} = (\mathcal{E}_k \setminus \partial_k(k)) \cup (U_k \setminus \{k\})$ (skip hyperedges $\partial_k(k)$, add hyperedge $U_k \setminus \{k\}$)

$\Phi_{k-1}(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_{k-1}}^{(k-1)} \bigotimes_{S \in \mathcal{E}_{k-1}} \psi_S(\mathbf{x}_S)$ (uses the new factor $\psi_{U_k \setminus \{k\}}$)

return $(\Phi_{k-1}, \mathcal{H}_{k-1} = (\mathcal{V}_{k-1}, \mathcal{E}_{k-1}))$

Procedure for Product Marginalisation

Procedure: ProductMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

Procedure for Product Marginalisation

Procedure: ProductMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

$\partial_k(k) = \{S \mid S \in \mathcal{E}_k \text{ with } k \in S\}$

Procedure for Product Marginalisation

Procedure: ProductMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

$\partial_k(k) = \{S \mid S \in \mathcal{E}_k \text{ with } k \in S\}$

for each $S \in \partial_k(k)$ marginalisation

$$\psi_{S\setminus\{k\}}(\mathbf{x}_{S\setminus\{k\}}) = \bigotimes_{x_k} \psi_S(\mathbf{x}_S)$$

for each $S \in \mathcal{E}_k \setminus \partial_k(k)$

$$\psi_S(\mathbf{x}_S) = \left(\psi_S(\mathbf{x}_S)\right)^{|\mathrm{Dom}(X_k)|}$$

Procedure for Product Marginalisation

Procedure: ProductMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

$\partial_k(k) = \{S \mid S \in \mathcal{E}_k \text{ with } k \in S\}$

for each $S \in \partial_k(k)$ marginalisation

$$\psi_{S\setminus\{k\}}(\mathbf{x}_{S\setminus\{k\}}) = \bigotimes_{x_k} \psi_S(\mathbf{x}_S)$$

for each $S \in \mathcal{E}_k \setminus \partial_k(k)$

$$\psi_S(\mathbf{x}_S) = \left(\psi_S(\mathbf{x}_S)\right)^{|\mathrm{Dom}(X_k)|}$$

$\mathcal{V}_{k-1} = [k-1]$

$\mathcal{E}_{k-1} = \{S \setminus \{k\} \mid S \in \mathcal{E}_k\}$

$\Phi_{k-1}(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_{k-1}}^{(k-1)} \bigotimes_{S \in \mathcal{E}_{k-1}} \psi_S(\mathbf{x}_S)$ (uses the new factors $\psi_{S\setminus\{k\}}, \psi_S$)

Procedure for Product Marginalisation

Procedure: ProductMarginalisation

Input:

- FAQ $\Phi_k(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_k}^{(k)} \bigotimes_{S \in \mathcal{E}_k} \psi_S(\mathbf{x}_S)$
- Hypergraph $\mathcal{H}_k = (\mathcal{V}_k, \mathcal{E}_k)$

$\partial_k(k) = \{S \mid S \in \mathcal{E}_k \text{ with } k \in S\}$

marginalisation

for each $S \in \partial_k(k)$

$$\psi_{S \setminus \{k\}}(\mathbf{x}_{S \setminus \{k\}}) = \bigotimes_{x_k} \psi_S(\mathbf{x}_S)$$

for each $S \in \mathcal{E}_k \setminus \partial_k(k)$

$$\psi_S(\mathbf{x}_S) = \left(\psi_S(\mathbf{x}_S)\right)^{|\mathrm{Dom}(X_k)|}$$

$\mathcal{V}_{k-1} = [k-1]$

$\mathcal{E}_{k-1} = \{S \setminus \{k\} \mid S \in \mathcal{E}_k\}$

$\Phi_{k-1}(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_{k-1}}^{(k-1)} \bigotimes_{S \in \mathcal{E}_{k-1}} \psi_S(\mathbf{x}_S)$ (uses the new factors $\psi_{S \setminus \{k\}}, \psi_S$)

return $(\Phi_{k-1}, \mathcal{H}_{k-1} = (\mathcal{V}_{k-1}, \mathcal{E}_{k-1}))$

Runtime of InsideOut

Consider an FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$ with hypergraph $\mathcal{H}$

Runtime of InsideOut

Consider an FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$ with hypergraph $\mathcal{H}$

Let $\mathcal{K} := \{k \mid k > f \text{ and } \bigoplus^{(k)} \neq \bigotimes\}$

Runtime of InsideOut

Consider an FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$ with hypergraph $\mathcal{H}$

Let $\mathcal{K} := \{k \mid k > f \text{ and } \bigoplus^{(k)} \neq \bigotimes\}$

Given the marginalisation order $\sigma = (f+1, \ldots, n)$, InsideOut constructs

- a sequence $\mathcal{H} = \mathcal{H}_n, \ldots, \mathcal{H}_f$ of hypergraphs
- a set U_k for each $k \in \mathcal{K}$
- a hypertree decomposition $\mathcal{T}$ for $\mathcal{H}_f$ with bags $(B_i)_{i \in [m]}$

Runtime of InsideOut

Consider an FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$ with hypergraph $\mathcal{H}$

Let $\mathcal{K} := \{k \mid k > f \text{ and } \bigoplus^{(k)} \neq \bigotimes\}$

Given the marginalisation order $\sigma = (f+1, \ldots, n)$, InsideOut constructs

- a sequence $\mathcal{H} = \mathcal{H}_n, \ldots, \mathcal{H}_f$ of hypergraphs
- a set U_k for each $k \in \mathcal{K}$
- a hypertree decomposition $\mathcal{T}$ for $\mathcal{H}_f$ with bags $(B_i)_{i \in [m]}$

The InsideOut Algorithm runs in time

$$\mathcal{O}\Big(\sum_{k \in \mathcal{K}} N^{\rho^*_{\mathcal{H}}(U_k)} + \sum_{i \in [m]} N^{\rho^*_{\mathcal{H}}(B_i)} + \text{OUT} \Big)$$

$N = |\max_{S \in \mathcal{E}} \psi_S|$

Runtime of InsideOut

Consider an FAQ $\Phi(\mathbf{x}_{[f]}) = \bigoplus_{x_{f+1}}^{(f+1)} \cdots \bigoplus_{x_n}^{(n)} \bigotimes_{S \in \mathcal{E}} \psi_S(\mathbf{x}_S)$ with hypergraph $\mathcal{H}$

Let $\mathcal{K} := \{k \mid k > f \text{ and } \bigoplus^{(k)} \neq \bigotimes\}$

Given the marginalisation order $\sigma = (f+1, \ldots, n)$, InsideOut constructs

- a sequence $\mathcal{H} = \mathcal{H}_n, \ldots, \mathcal{H}_f$ of hypergraphs
- a set U_k for each $k \in \mathcal{K}$
- a hypertree decomposition $\mathcal{T}$ for $\mathcal{H}_f$ with bags $(B_i)_{i \in [m]}$

The InsideOut Algorithm runs in time

$$\mathcal{O}\Big(\sum_{k \in \mathcal{K}} N^{\rho^*_{\mathcal{H}}(U_k)} + \sum_{i \in [m]} N^{\rho^*_{\mathcal{H}}(B_i)} + \text{OUT}\Big) = \mathcal{O}\Big(N^{\text{faqw}(\sigma,\mathcal{T})} + \text{OUT}\Big)$$

$N = |\max_{S \in \mathcal{E}} \psi_S|$

$\text{faqw}(\sigma, \mathcal{T}) = \max_{k \in \mathcal{K}, i \in [m]} \{\rho^*_{\mathcal{H}}(U_k), \rho^*_{\mathcal{H}}(B_i)\}$ is the FAQ-width of $(\sigma, \mathcal{T})$

FAQ-Width of a Query

Consider an FAQ Φ

FAQ-Width of a Query

Consider an FAQ Φ

- Let $\mathrm{VMO}(\Phi)$ be the set of valid marginalisation orders for Φ

FAQ-Width of a Query

Consider an FAQ Φ

- Let $\text{VMO}(\Phi)$ be the set of valid marginalisation orders for Φ
- For $\sigma \in \text{VMO}(\Phi)$, let $\mathbf{T}(\sigma)$ be the set of hypertree decompositions for the query obtained from Φ after marginalising the bound variables following σ

FAQ-Width of a Query

Consider an FAQ Φ

- Let VMO(Φ) be the set of valid marginalisation orders for Φ
- For $\sigma \in$ VMO(Φ), let $\mathbf{T}(\sigma)$ be the set of hypertree decompositions for the query obtained from Φ after marginalising the bound variables following σ

FAQ width of Φ: $\text{faqw}(\Phi) = \min_{\sigma \in \text{VMO}(\Phi), \mathcal{T} \in \mathbf{T}(\sigma)} \text{faqw}(\sigma, \mathcal{T})$

FAQ-Width of a Query

Consider an FAQ Φ

- Let $\mathsf{VMO}(\Phi)$ be the set of valid marginalisation orders for Φ
- For $\sigma \in \mathsf{VMO}(\Phi)$, let $\mathbf{T}(\sigma)$ be the set of hypertree decompositions for the query obtained from Φ after marginalising the bound variables following σ

FAQ width of Φ: $\mathsf{faqw}(\Phi) = \min_{\sigma \in \mathsf{VMO}(\Phi), \mathcal{T} \in \mathbf{T}(\sigma)} \mathsf{faqw}(\sigma, \mathcal{T})$

For optimal σ and $\mathcal{T}$, InsideOut computes Φ in time

$$\mathcal{O}\left(N^{\mathsf{faqw}(\Phi)} + \mathsf{OUT}\right)$$