Optimal tensor-matrix and -vector contraction strategies for computing dense tensor decompositions

Oguz Kaya

Université Paris-Sud/Paris-Saclay and LRI, Orsay, France

June 28, 2019

Joint work with:

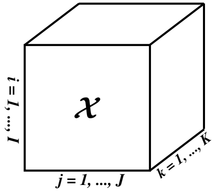
Yves Robert, LIP and ENS Lyon, France

A tensor?

What is a tensor?

- A tensor is a **multi-dimensional array**.
  - A vector is a 1-dimensional tensor.
  - A matrix is a 2-dimensional tensor.
  - A tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_D}$ has D dimensions.
- A tensor is a **finite, discrete, multi-variate function**.
  - $\mathcal{X} : \mathcal{I}_1 \times \cdots \times \mathcal{I}_D \to \mathbb{R}, \mathcal{I}_d = \{1, 2, \ldots, I_d\}$
  - Example: $\mathcal{X}(i_1, \ldots, i_D) = k,\ k \in \mathbb{R}$

We are interested in the case when $\mathcal{X}$ is of **low rank**.



$\mathcal{X} \in \mathbb{R}^{I \times J \times K}$

Outline

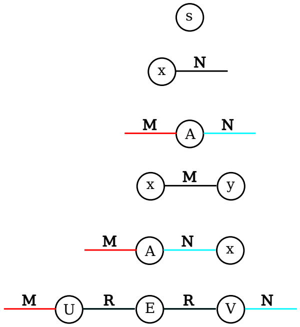
1 Introduction

2 Tensor networks/decompositions/contraction

3 Optimal contractions using dimension trees

4 Conclusion

Tensor networks



scalar

vector

matrix

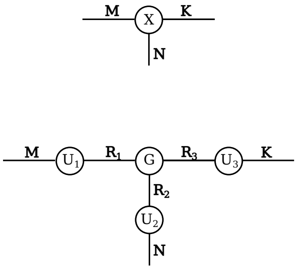
$x^T y$

Ax

$A=UEV^T$

🟥 red=row

🟦 cyan=col

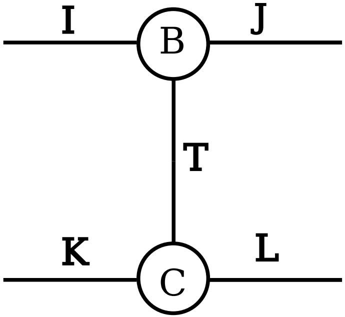


full tensor

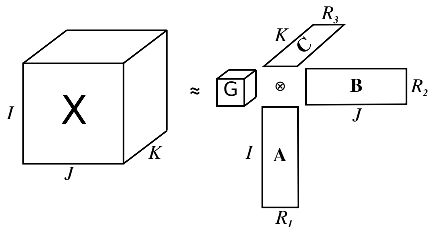
tensor in
Tucker form
R << M,N,K

Tensor contraction

- $\mathcal{A}("ijkl") = \mathcal{B}("ijt") * \mathcal{C}("tkl")$
- $\mathcal{A}(i,j,k,l) = \sum_t \mathcal{B}(i,j,t)\mathcal{C}(t,k,l)$
- **Cost:** $O(IJTKL)$



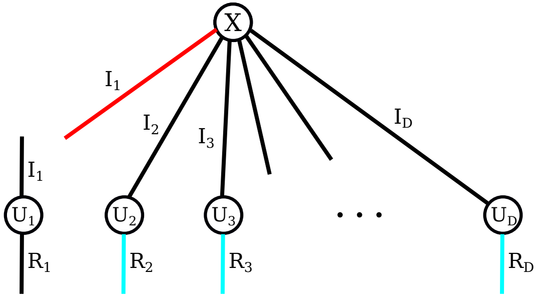
Tucker decomposition and its computation



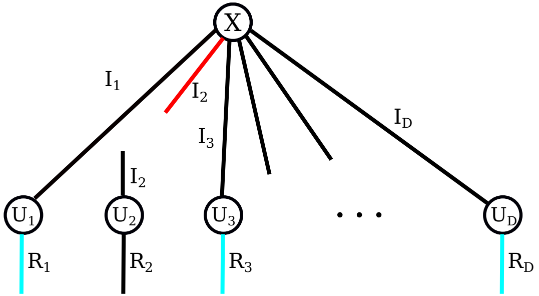
Algorithm HOOI for 3rd order tensors

```
Input: $\mathcal{X}$: Tensor
  $(R_1, R_2, R_3)$: The rank of approximation
Output: Tucker decomposition $[\![\mathcal{G}; \mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$
  repeat
    $\hat{\mathbf{A}} \leftarrow (\mathcal{X} \times_2 \mathbf{B} \times_3 \mathbf{C})_{(1)}$
    $\mathbf{A} \leftarrow \mathrm{TRSVD}(\hat{\mathbf{A}}, R_1)$
    $\hat{\mathbf{B}} \leftarrow (\mathcal{X} \times_1 \mathbf{A} \times_3 \mathbf{C})_{(2)}$
    $\mathbf{B} \leftarrow \mathrm{TRSVD}(\hat{\mathbf{B}}, R_2)$
    $\hat{\mathbf{C}} \leftarrow (\mathcal{X} \times_1 \mathbf{A} \times_2 \mathbf{B})_{(3)}$
    $\mathbf{C} \leftarrow \mathrm{TRSVD}(\hat{\mathbf{C}}, R_3)$
  until no improvement or max iterations achieved
  $\mathcal{G} \leftarrow \mathcal{X} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C}$
```

HOOI and contractions



HOOI and contractions



Outline

1 Introduction

2 Tensor networks/decompositions/contraction

3 Optimal contractions using dimension trees

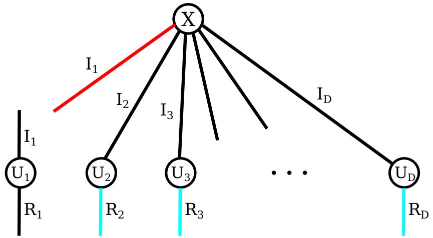
4 Conclusion

Optimal tensor-matrix contractions

Theorem

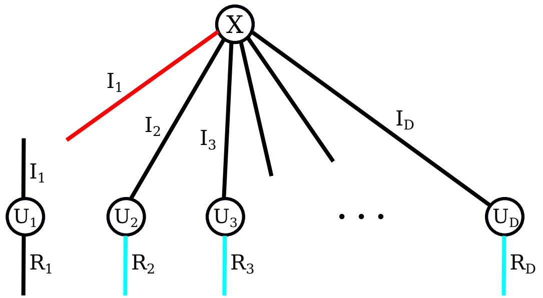
Given a tensor $\mathcal{X} \in \mathbb{R}^{I_1 \times \cdots \times I_D}$ and the set of matrices $\{\mathbf{U}_d \in \mathbb{R}^{I_d \times R_d}\}$, an optimal contraction sequence involving $\mathcal{X}$ and matrices $\mathbf{U}_i$ and $\mathbf{U}_j$ iff $R_i/I_i \leq R_j/I_j$.

- Contractions are commutative.
- NP-hard for arbitrary tensor networks.
- $O(D \log D)$ optimal greedy algorithm for finding optimal contraction sequence.



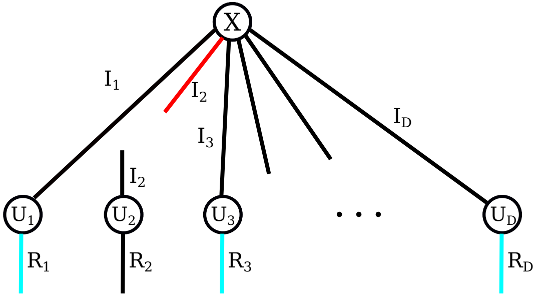
Reusing contractions in HOOI

- Contractions are commutative.



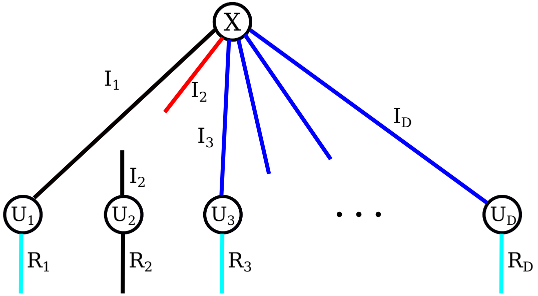
Reusing contractions in HOOI

- Contractions are commutative.



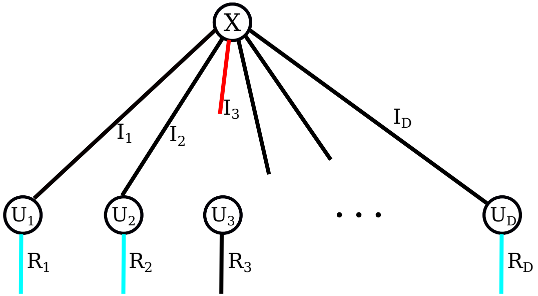
Reusing contractions in HOOI

- Contractions are commutative.

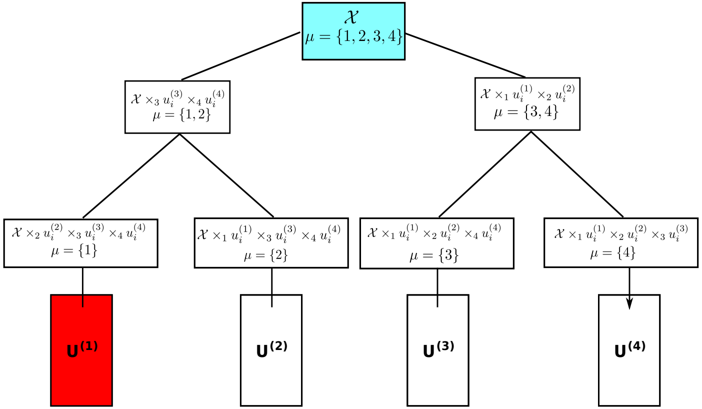


Reusing contractions in HOOI

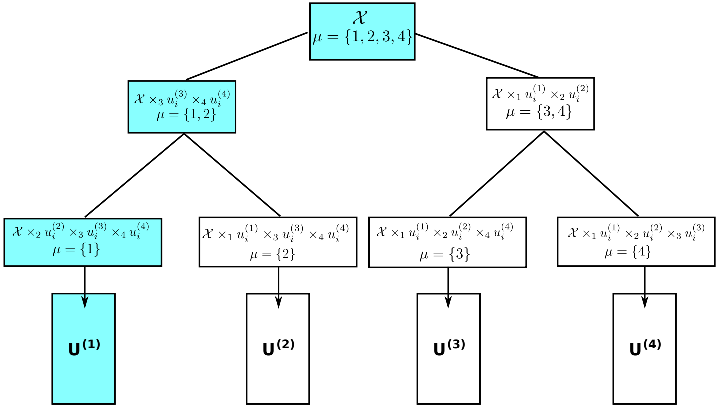
- Contractions are commutative.



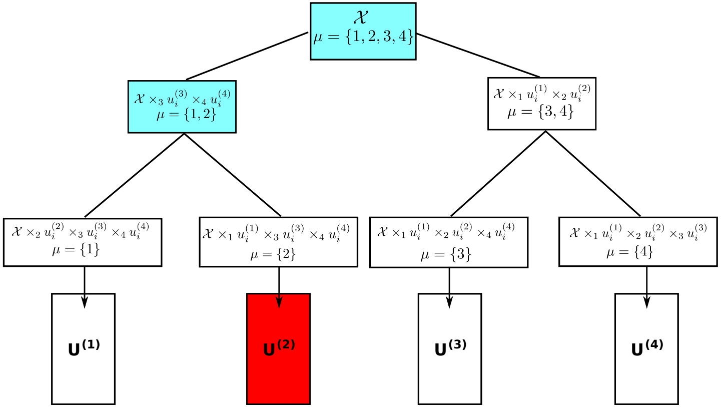
HOOI contractions using dimension tree


$\mathcal{X}$ $\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$ $\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$ $\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$ $\mu = \{4\}$
$\mathbf{U}^{(1)}$
$\mathbf{U}^{(2)}$
$\mathbf{U}^{(3)}$
$\mathbf{U}^{(4)}$

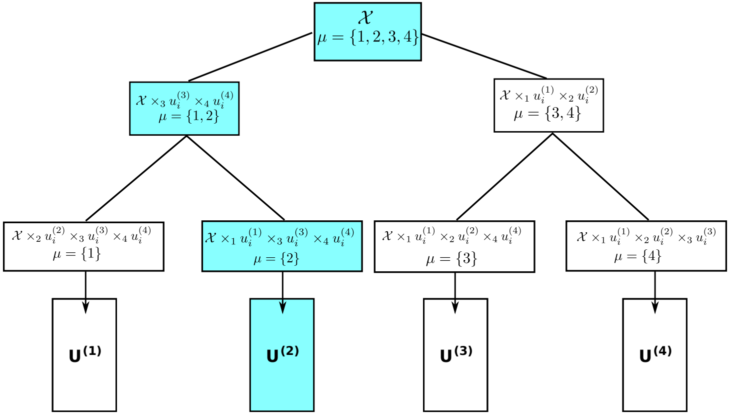
HOOI contractions using dimension tree


$\mathcal{X}$
$\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$
$\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$
$\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$
$\mu = \{4\}$
$\mathbf{U^{(1)}}$
$\mathbf{U^{(2)}}$
$\mathbf{U^{(3)}}$
$\mathbf{U^{(4)}}$

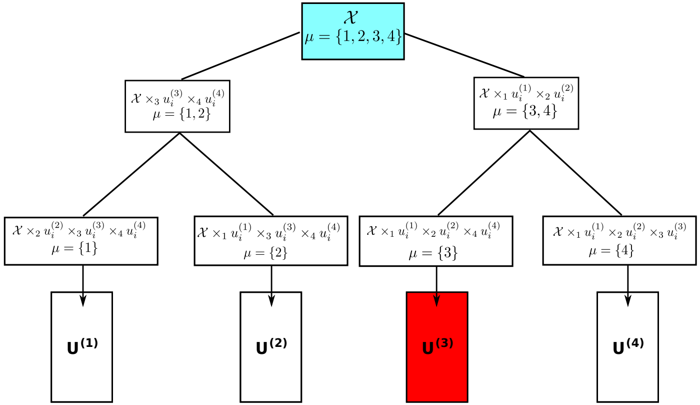
HOOI contractions using dimension tree


$\mathcal{X}$
$\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$
$\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$
$\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$
$\mu = \{4\}$
$\mathbf{U}^{(1)}$
$\mathbf{U}^{(2)}$
$\mathbf{U}^{(3)}$
$\mathbf{U}^{(4)}$

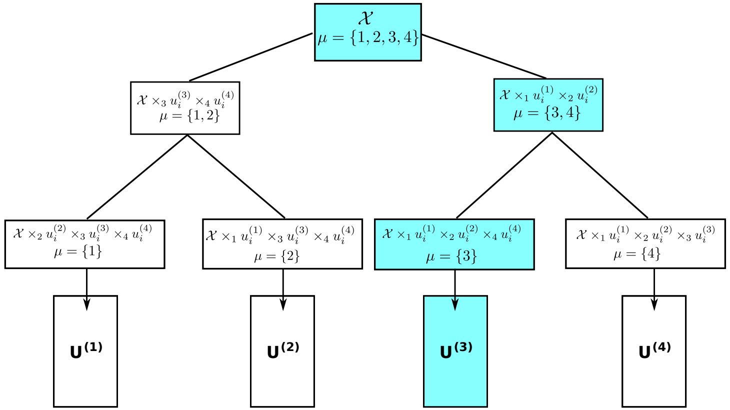
HOOI contractions using dimension tree


$\mathcal{X}$
$\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$
$\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$
$\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$
$\mu = \{4\}$
$\mathbf{U}^{(1)}$
$\mathbf{U}^{(2)}$
$\mathbf{U}^{(3)}$
$\mathbf{U}^{(4)}$

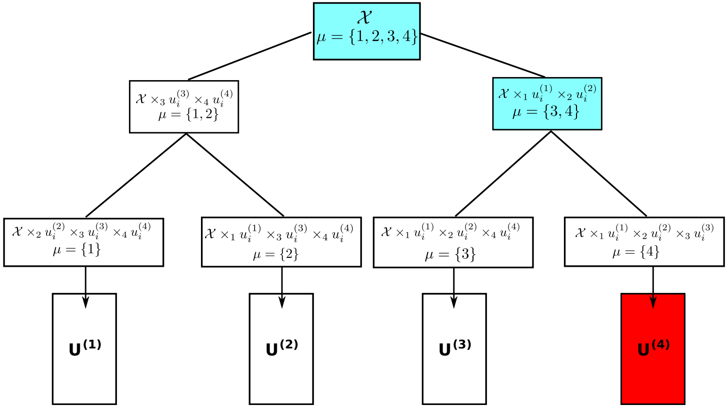
HOOI contractions using dimension tree


$\mathcal{X}$
$\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$
$\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$
$\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$
$\mu = \{4\}$
$\mathbf{U^{(1)}}$
$\mathbf{U^{(2)}}$
$\mathbf{U^{(3)}}$
$\mathbf{U^{(4)}}$

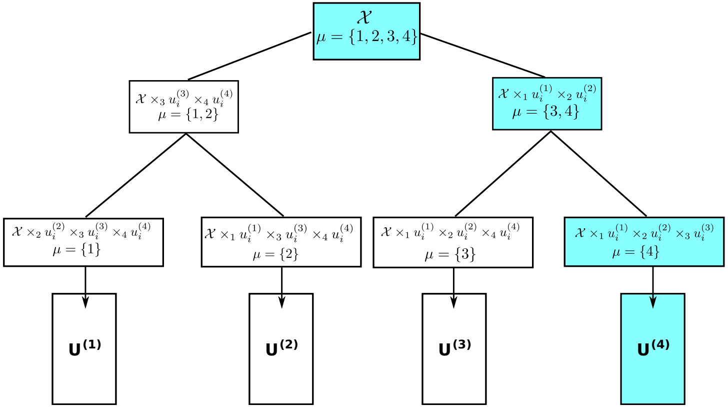
HOOI contractions using dimension tree


$\mathcal{X}$
$\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$
$\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$
$\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$
$\mu = \{4\}$
$\mathbf{U}^{(1)}$
$\mathbf{U}^{(2)}$
$\mathbf{U}^{(3)}$
$\mathbf{U}^{(4)}$

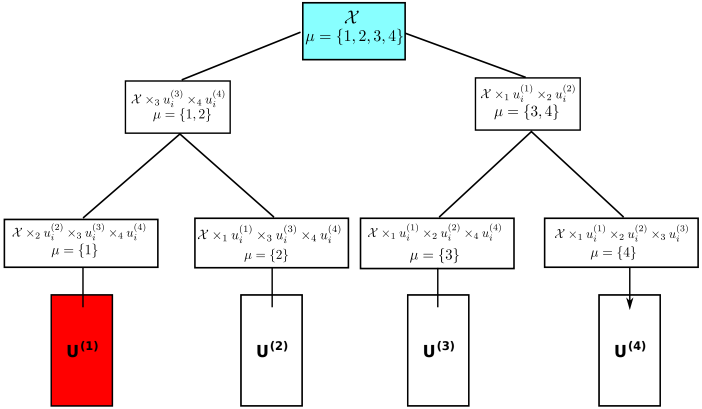
HOOI contractions using dimension tree


$\mathcal{X}$ $\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$ $\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$ $\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$ $\mu = \{4\}$
$\mathbf{U^{(1)}}$
$\mathbf{U^{(2)}}$
$\mathbf{U^{(3)}}$
$\mathbf{U^{(4)}}$

HOOI contractions using dimension tree

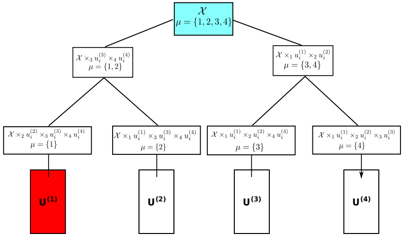

$\mathcal{X}$
$\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$
$\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$
$\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$
$\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$
$\mu = \{4\}$
$\mathbf{U}^{(1)}$
$\mathbf{U}^{(2)}$
$\mathbf{U}^{(3)}$
$\mathbf{U}^{(4)}$

HOOI contractions using dimension tree


$\mathcal{X}$ $\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$ $\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$ $\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$ $\mu = \{4\}$
$\mathbf{U}^{(1)}$
$\mathbf{U}^{(2)}$
$\mathbf{U}^{(3)}$
$\mathbf{U}^{(4)}$

Dimension tree-based computation

- Each node involves an index set corresponding to non-contracted dimensions.
- Index sets of children **partition** that of their parent.
- Contraction of a child node can always be computed using its parent.
- Leaves have single dimension index.
- $O(D \log D)$ contractions per iteration (instead of $O(D^2)$).



$\mathcal{X}$ $\mu = \{1,2,3,4\}$
$\mathcal{X} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{1,2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)}$ $\mu = \{3,4\}$
$\mathcal{X} \times_2 u_i^{(2)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{1\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_3 u_i^{(3)} \times_4 u_i^{(4)}$ $\mu = \{2\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_4 u_i^{(4)}$ $\mu = \{3\}$
$\mathcal{X} \times_1 u_i^{(1)} \times_2 u_i^{(2)} \times_3 u_i^{(3)}$ $\mu = \{4\}$
$\mathbf{U}^{(1)}$
$\mathbf{U}^{(2)}$
$\mathbf{U}^{(3)}$
$\mathbf{U}^{(4)}$

Optimal dimension tree - Structure?

Is optimal tree necessarily binary? Balanced?

Optimal dimension tree - Structure?

Is optimal tree necessarily binary? Balanced?

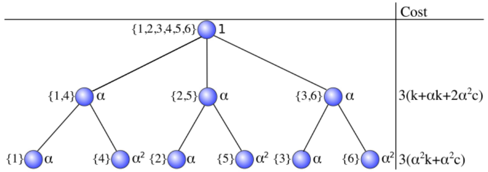


Fig. 3 A ternary dimension tree for $\mathcal{X}$. The μ set of the node is provided on the left, and the tensor size coefficient is provided on the right of each tree node. The TTM cost of each level is provided on the right

Optimal dimension tree - Structure?

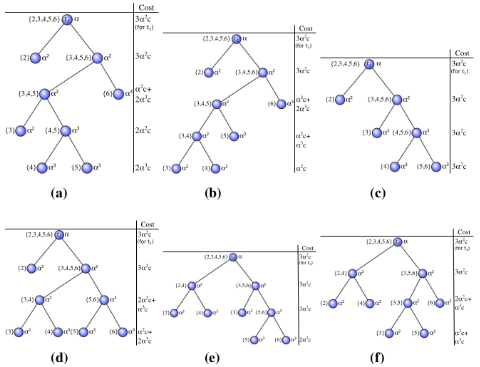


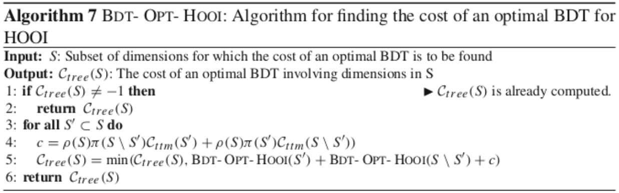
Fig. 4 All possible configurations and associated TTM costs at each level of BDTs rooted at t_2. $\mu(t)$ is given on the left of each tree node t, On the right of a tree node, the size coefficient of its tensor is provided, i.e., t_2 has a tensor of size $(I_1 I_2 I_3 I_4 I_5 I_6)\alpha$

Optimal dimension tree - Complexity?

Theorem

Finding an optimal binary dimension tree for HOOI contractions is NP-hard.

Optimal dimension tree - Algorithm?



Algorithm 7 BDT- OPT- HOOI: Algorithm for finding the cost of an optimal BDT for HOOI

Input: S: Subset of dimensions for which the cost of an optimal BDT is to be found

Output: $\mathcal{C}_{tree}(S)$: The cost of an optimal BDT involving dimensions in S

```
1: if $\mathcal{C}_{tree}(S) \neq -1$ then                    ▶ $\mathcal{C}_{tree}(S)$ is already computed.
2:     return $\mathcal{C}_{tree}(S)$
3: for all $S' \subset S$ do
4:     $c = \rho(S)\pi(S \setminus S')\mathcal{C}_{ttm}(S') + \rho(S)\pi(S')\mathcal{C}_{ttm}(S \setminus S')$
5:     $\mathcal{C}_{tree}(S) = \min(\mathcal{C}_{tree}(S)$, BDT- OPT- HOOI$(S')$ + BDT- OPT- HOOI$(S \setminus S') + c)$
6: return $\mathcal{C}_{tree}(S)$
```

- **Computational cost:** $O(3^D)$
- **Memory cost:** $O(2^D)$
- $(O((k+1)^D))$ for k-ary trees)

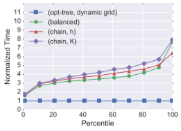
Optimal dimension tree for CP decomposition

- Optimal tree **is** binary.
- Finding the optimal tree is still NP-hard.
- Optimal binary tree found in $O(3^D)$ time and $O(2^D)$ space.

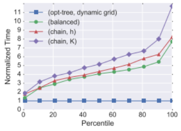
Some results (from Chakaravarthy et al.)

	Tensor Dimensions	Core Tensor Dimensions
HCCI	(672, 672, 627, 16)	(279, 279, 153, 14)
TJLR	(460, 700, 360, 16, 4)	(306, 232, 239, 16, 4)
SP	(500, 500, 500, 11, 10)	(81, 129, 127, 7, 6)

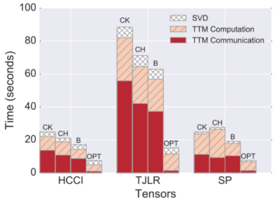
Table 2: Real tensors used in our study



(a) Overall time (5D)



(b) Overall time (6D)



(c) Real Tensors. CK:(chain,K), CH: (chain, h), B: (balanced), OPT:(opt-tree, dynamic grid)

Figure 10: Overall Execution Time

Outline

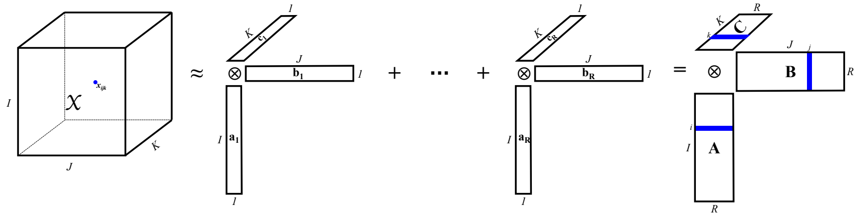
1 Introduction

2 Tensor networks/decompositions/contraction

3 Optimal contractions using dimension trees

4 Conclusion

Applications



- Recommender systems, web search, link prediction, healthcare, etc.
- Numerical data compression (Kolda et al., '16)
- Signal processing (blind source separation)
- Scientific computing (quantum chemistry, hierarchical low-rank methods, PDEs, etc.)
- Deep learning

References

O. Kaya and Y. Robert. Computing dense tensor decompositions with optimal dimension trees. *Algorithmica*, 2019.

O. Kaya and B. Uçar. Parallel computation of CANDECOMP/PARAFAC decomposition using dimension trees. *SIAM Journal on Scientific Computing*, 2018.

V. Chakaravathy, J. Choi, D. Joseph, P. Murali, S. Pandian, Y. Sabharwal, D. Sreedhar. On Optimizing Distributed Tucker Decomposition for Sparse Tensors. *ICS'18*, 2018.

Contact

Oguz Kaya
Université Paris-Sud/Paris-Saclay and LRI, Orsay, France
oguz.kaya@lri.fr
www.oguzkaya.com