Fast learning of fast transforms, with guarantees

Ecole d'été de Peyresq 2022

Quoc-Tung Le, Léon Zheng, Elisa Riccietti, Rémi Gribonval

June, 2022

Approximating a matrix by a product of sparse factors

Given a matrix $\mathbf{Z}$ and $J \geq 2$, find sparse factors $\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}$ such that

$$\mathbf{Z} \approx \mathbf{X}^{(J)}\mathbf{X}^{(J-1)} \dots \mathbf{X}^{(1)}.$$

Approximating a matrix by a product of sparse factors

Given a matrix $\mathbf{Z}$ and $J \geq 2$, find sparse factors $\mathbf{X}^{(J)}, \ldots, \mathbf{X}^{(1)}$ such that

$$\mathbf{Z} \approx \mathbf{X}^{(J)} \mathbf{X}^{(J-1)} \ldots \mathbf{X}^{(1)}.$$

Application (Large-scale inverse problem)

Reduce time/memory complexity: find $\mathbf{x}$ such that $\mathbf{y} = \mathbf{Z}\mathbf{x}$.

Approximating a matrix by a product of sparse factors

Given a matrix $\mathbf{Z}$ and $J \geq 2$, find sparse factors $\mathbf{X}^{(J)}, \ldots, \mathbf{X}^{(1)}$ such that

$$\mathbf{Z} \approx \mathbf{X}^{(J)} \mathbf{X}^{(J-1)} \ldots \mathbf{X}^{(1)}.$$

Application (Large-scale inverse problem)

Reduce time/memory complexity: find $\mathbf{x}$ such that $\mathbf{y} = (\mathbf{X}^{(J)} \ldots \mathbf{X}^{(1)})\mathbf{x}$.

Approximating a matrix by a product of sparse factors

Given a matrix $\mathbf{Z}$ and $J \geq 2$, find sparse factors $\mathbf{X}^{(J)}, \ldots, \mathbf{X}^{(1)}$ such that

$$\mathbf{Z} \approx \mathbf{X}^{(J)}\mathbf{X}^{(J-1)} \ldots \mathbf{X}^{(1)}.$$

Application (Large-scale inverse problem)

Reduce time/memory complexity: find $\mathbf{x}$ such that $\mathbf{y} = (\mathbf{X}^{(J)} \ldots \mathbf{X}^{(1)})\mathbf{x}$.

Application (Deep neural network compression)

Replace a dense **weight** matrix $\mathbf{Z}$ by $\mathbf{X}^{(J)} \ldots \mathbf{X}^{(1)}$.

Figure: Variants of Vision Transformers [Dosovitskiy et al, 2021]

Model	Layers	Hidden size D	MLP size	Heads	Params
ViT-Base	12	768	3072	12	86M
ViT-Large	24	1024	4096	16	307M
ViT-Huge	32	1280	5120	16	632M

Approximating a matrix by a product of sparse factors

Problem formulation

$$\min_{\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(J)}} \left\| \mathbf{Z} - \mathbf{X}^{(J)}\mathbf{X}^{(J-1)}\ldots\mathbf{X}^{(1)} \right\|_F^2, \quad \text{such that } \{\mathbf{X}^{(\ell)}\}_\ell \text{ are sparse.} \quad (1)$$

Choices for sparsity constraint:

1. **Classical sparsity patterns**: k-sparsity by column and/or by row
2. **Fixed-support constraint**: $\text{supp}(\mathbf{X}^{(\ell)}) \subseteq \mathbf{S}^{(\ell)}$ for $\ell = 1, \ldots, J$.

Approximating a matrix by a product of sparse factors

Problem formulation

$$\min_{\mathbf{X}^{(1)},\ldots,\mathbf{X}^{(J)}} \left\| \mathbf{Z} - \mathbf{X}^{(J)}\mathbf{X}^{(J-1)}\ldots\mathbf{X}^{(1)} \right\|_F^2, \quad \text{such that } \{\mathbf{X}^{(\ell)}\}_\ell \text{ are sparse.} \quad (1)$$

Choices for sparsity constraint:

1. **Classical sparsity patterns**: k-sparsity by column and/or by row
2. **Fixed-support constraint**: $\text{supp}(\mathbf{X}^{(\ell)}) \subseteq \mathbf{S}^{(\ell)}$ for $\ell = 1, \ldots, J$.

A difficult problem

- Sparse coding is NP-hard [Foucart et al. 2013].
- Fixed-support setting is NP-hard for $J = 2$ factors [Le et al. 2021].
- Gradient-based methods [Le Magoarou et al. 2016] lack guarantees.

Focus on fixed-support constraint

When is the problem well-posed and tractable? (case with $J = 2$)

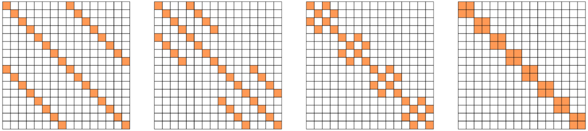
1. Conditions for uniqueness of the solution [Zheng et al. 2022]
2. Conditions for achieving global optimality [Le et al. 2021]

Focus on fixed-support constraint

When is the problem well-posed and tractable? (case with $J = 2$)

1. Conditions for uniqueness of the solution [Zheng et al. 2022]
2. Conditions for achieving global optimality [Le et al. 2021]

→ We study a fixed-support constraint ($J \geq 2$) satisfying such conditions.



(a) $\mathbf{S}_{\mathrm{bf}}^{(4)}$ (b) $\mathbf{S}_{\mathrm{bf}}^{(3)}$ (c) $\mathbf{S}_{\mathrm{bf}}^{(2)}$ (d) $\mathbf{S}_{\mathrm{bf}}^{(1)}$

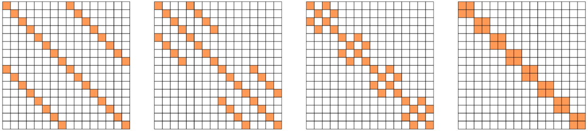
Figure: Butterfly structure: $\mathrm{supp}(\mathbf{X}^{(\ell)}) \subseteq \mathbf{S}_{\mathrm{bf}}^{(\ell)} := \mathbf{I}_{N/2^\ell} \otimes \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right] \otimes \mathbf{I}_{2^{\ell-1}}$.

Focus on fixed-support constraint

When is the problem well-posed and tractable? (case with $J = 2$)

1. Conditions for uniqueness of the solution [Zheng et al. 2022]
2. Conditions for achieving global optimality [Le et al. 2021]

$\rightarrow$ We study a fixed-support constraint ($J \geq 2$) satisfying such conditions.



(a) $\mathbf{S}_{\text{bf}}^{(4)}$ (b) $\mathbf{S}_{\text{bf}}^{(3)}$ (c) $\mathbf{S}_{\text{bf}}^{(2)}$ (d) $\mathbf{S}_{\text{bf}}^{(1)}$

Figure: Butterfly structure: $\text{supp}(\mathbf{X}^{(\ell)}) \subseteq \mathbf{S}_{\text{bf}}^{(\ell)} := \mathbf{I}_{\mathbf{N}/\mathbf{2}^{\ell}} \otimes \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right] \otimes \mathbf{I}_{\mathbf{2}^{\ell-1}}$.

Why butterfly structure?

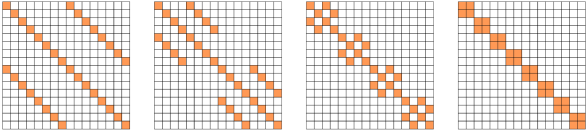
- Allows fast $\mathcal{O}(N \log N)$ matrix-vector multiplication
- Captures DFT, DCT, Hadamard, convolution, ... [Dao et al. 2020]

Focus on fixed-support constraint

When is the problem well-posed and tractable? (case with $J = 2$)

1. Conditions for uniqueness of the solution [Zheng et al. 2022]
2. Conditions for achieving global optimality [Le et al. 2021]

$\rightarrow$ We study a fixed-support constraint ($J \geq 2$) satisfying such conditions.



(a) $\mathbf{S}_{\text{bf}}^{(4)}$ (b) $\mathbf{S}_{\text{bf}}^{(3)}$ (c) $\mathbf{S}_{\text{bf}}^{(2)}$ (d) $\mathbf{S}_{\text{bf}}^{(1)}$

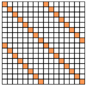
Figure: Butterfly structure: $\text{supp}(\mathbf{X}^{(\ell)}) \subseteq \mathbf{S}_{\text{bf}}^{(\ell)} := \mathbf{I}_{\mathbf{N}/\mathbf{2}^{\ell}} \otimes \left[\begin{smallmatrix} 1 & 1 \\ 1 & 1 \end{smallmatrix}\right] \otimes \mathbf{I}_{\mathbf{2}^{\ell-1}}$.

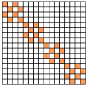
Main contribution

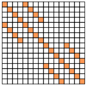
An efficient **hierarchical algorithm** to approximate **any** matrix by a product of butterfly factors.

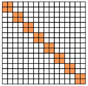
Hierarchical factorization algorithm

Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$ 

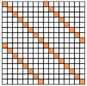
$\mathrm{supp}(\mathbf{X}^{(3)}) \subseteq$ 

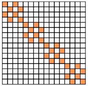
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$ 

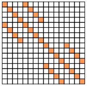
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$ 

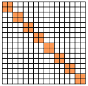
Hierarchical factorization algorithm

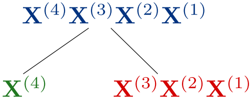
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$ 

$\mathrm{supp}(\mathbf{X}^{(3)}) \subseteq$ 

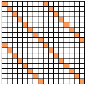
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$ 

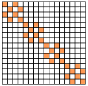
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$ 

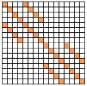


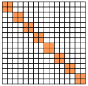
Hierarchical factorization algorithm

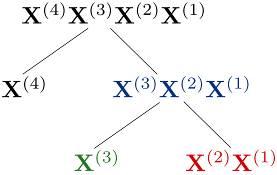
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$ 

$\mathrm{supp}(\mathbf{X}^{(3)}) \subseteq$ 

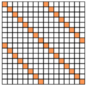
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$ 

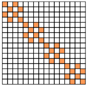
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$ 

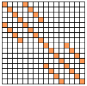


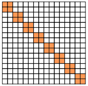
Hierarchical factorization algorithm

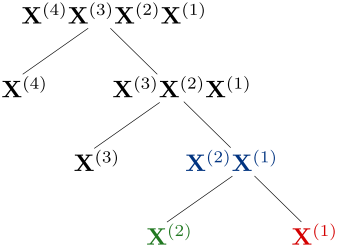
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$ 

$\mathrm{supp}(\mathbf{X}^{(3)}) \subseteq$ 

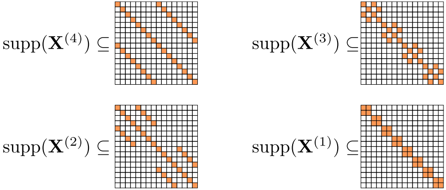
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$ 

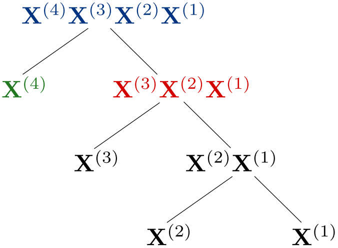
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$ 



Hierarchical factorization algorithm

Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

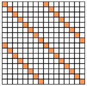


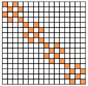


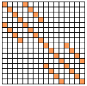
How to recover the partial products?

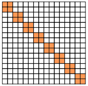
Hierarchical factorization algorithm

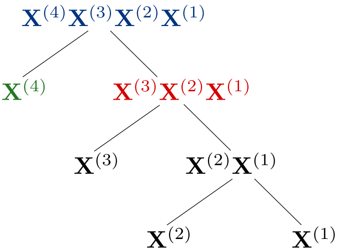
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

$\text{supp}(\mathbf{X}^{(4)}) \subseteq$ 

$\text{supp}(\mathbf{X}^{(3)}) \subseteq$ 

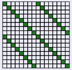
$\text{supp}(\mathbf{X}^{(2)}) \subseteq$ 

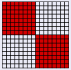
$\text{supp}(\mathbf{X}^{(1)}) \subseteq$ 



How to recover the partial products? $\rightarrow$ use their known supports

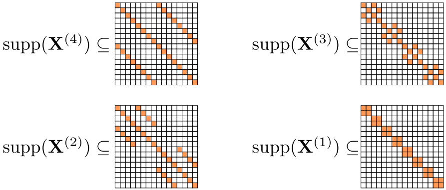
Lemma (Supports of the partial products)

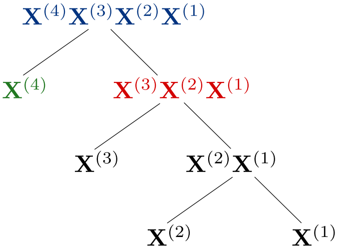
$\text{supp}(\mathbf{X}^{(4)}) \subseteq$  $= \mathbf{S}_{\texttt{bf}}^{(4)}$

$\text{supp}(\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}) \subseteq$  $= \mathbf{S}_{\texttt{bf}}^{(3)}\mathbf{S}_{\texttt{bf}}^{(2)}\mathbf{S}_{\texttt{bf}}^{(1)}$

Hierarchical factorization algorithm

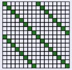
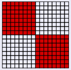
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:





How to recover the partial products? $\rightarrow$ use their known supports

Lemma (Supports of the partial products)

$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(4)}$ $\qquad$ $\mathrm{supp}(\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)}$

Two-layer fixed-support problem:

$$\min_{\mathbf{A},\mathbf{B}} \|\mathbf{Z} - \mathbf{A}\mathbf{B}\|_F^2, \ \text{s.t.}\ \mathrm{supp}(\mathbf{A}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(4)},\ \mathrm{supp}(\mathbf{B}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)} \qquad (2)$$

Two-layer fixed-support sparse matrix factorization

$$\min_{\mathbf{A},\mathbf{B}} \|\mathbf{Z} - \mathbf{A}\mathbf{B}\|_F^2 \,, \text{ s.t. } \operatorname{supp}(\mathbf{A}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(4)}, \ \operatorname{supp}(\mathbf{B}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)} \qquad (2)$$

Two-layer fixed-support sparse matrix factorization

$$\min_{\mathbf{A},\mathbf{B}} \|\mathbf{Z} - \mathbf{A}\mathbf{B}\|_F^2 \,, \text{ s.t. } \operatorname{supp}(\mathbf{A}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(4)}, \ \operatorname{supp}(\mathbf{B}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)} \qquad (2)$$

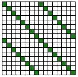
Fact: $\mathbf{AB} = \sum_{i=1}^{N} \mathbf{A}_{\bullet,i}\mathbf{B}_{i,\bullet}$.

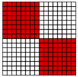
Two-layer fixed-support sparse matrix factorization

$$\min_{\mathbf{A},\mathbf{B}} \|\mathbf{Z} - \mathbf{A}\mathbf{B}\|_F^2 \,, \text{ s.t. } \operatorname{supp}(\mathbf{A}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(4)}, \ \operatorname{supp}(\mathbf{B}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)} \qquad (2)$$

Fact: $\mathbf{AB} = \sum_{i=1}^{N} \mathbf{A}_{\bullet,i}\mathbf{B}_{i,\bullet}$.

Constraint on the pair of factors

$\operatorname{supp}(\mathbf{A}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(4)}$

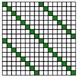
$\operatorname{supp}(\mathbf{B}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)}$

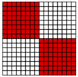
Two-layer fixed-support sparse matrix factorization

$$\min_{\mathbf{A},\mathbf{B}} \|\mathbf{Z} - \mathbf{A}\mathbf{B}\|_F^2, \text{ s.t. } \mathrm{supp}(\mathbf{A}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(4)}, \ \mathrm{supp}(\mathbf{B}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)} \quad (2)$$

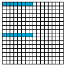
Fact: $\mathbf{AB} = \sum_{i=1}^{N} \mathbf{A}_{\bullet,i}\mathbf{B}_{i,\bullet}$.

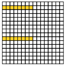
Constraint on the pair of factors

$\mathrm{supp}(\mathbf{A}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(4)}$

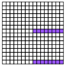
$\mathrm{supp}(\mathbf{B}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)}$

Constraint on the rank-one matrices

$\mathrm{supp}(\mathbf{A}_{\bullet,1}\mathbf{B}_{1,\bullet}) \subseteq$  $= \mathcal{S}_1$

$\mathrm{supp}(\mathbf{A}_{\bullet,2}\mathbf{B}_{2,\bullet}) \subseteq$  $= \mathcal{S}_2$

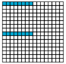
$\vdots$

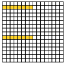
$\mathrm{supp}(\mathbf{A}_{\bullet,N}\mathbf{B}_{N,\bullet}) \subseteq$  $= \mathcal{S}_N$

Two-layer fixed-support sparse matrix factorization

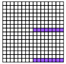
$$\min_{\mathbf{A},\mathbf{B}} \|\mathbf{Z}-\mathbf{A}\mathbf{B}\|_F^2\,, \text{ s.t. } \mathrm{supp}(\mathbf{A}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(4)},\ \mathrm{supp}(\mathbf{B}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)} \qquad (2)$$

Constraint on the rank-one matrices

$\mathrm{supp}(\mathbf{A}_{\bullet,1}\mathbf{B}_{1,\bullet}) \subseteq$  $= \mathcal{S}_1$

$\mathrm{supp}(\mathbf{A}_{\bullet,2}\mathbf{B}_{2,\bullet}) \subseteq$  $= \mathcal{S}_2$

$\cdots$

$\mathrm{supp}(\mathbf{A}_{\bullet,N}\mathbf{B}_{N,\bullet}) \subseteq$  $= \mathcal{S}_N$

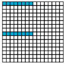
Theorem ([Le et al. 2021; Zheng et al. 2022])

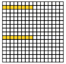
The rank-one matrices have **pairwise disjoint supports**. *Consequently, (1) is polynomially solvable and admits an essentially unique solution.*

Two-layer fixed-support sparse matrix factorization

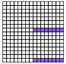
$$\min_{\mathbf{A},\mathbf{B}} \|\mathbf{Z} - \mathbf{A}\mathbf{B}\|_F^2 \,, \text{ s.t. } \mathrm{supp}(\mathbf{A}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(4)}, \ \mathrm{supp}(\mathbf{B}) \subseteq \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)} \qquad (2)$$

Constraint on the rank-one matrices

$\mathrm{supp}(\mathbf{A}_{\bullet,1}\mathbf{B}_{1,\bullet}) \subseteq$  $= \mathcal{S}_1$

$\mathrm{supp}(\mathbf{A}_{\bullet,2}\mathbf{B}_{2,\bullet}) \subseteq$  $= \mathcal{S}_2$

$\cdots$

$\mathrm{supp}(\mathbf{A}_{\bullet,N}\mathbf{B}_{N,\bullet}) \subseteq$  $= \mathcal{S}_N$

Theorem ([Le et al. 2021; Zheng et al. 2022])

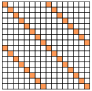
The rank-one matrices have **pairwise disjoint supports**. *Consequently, (1) is polynomially solvable and admits an essentially unique solution.*

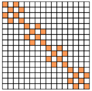
Algorithm to solve (1):

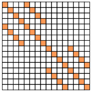
1. Extract the submatrices $\mathbf{Z}_{|\mathcal{S}_i}$, $i = 1, \ldots, N$
2. Perform best rank-one approximation for each submatrix

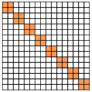
Hierarchical factorization algorithm

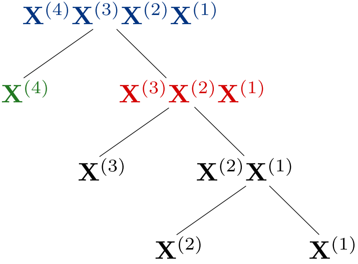
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$ 

$\mathrm{supp}(\mathbf{X}^{(3)}) \subseteq$ 

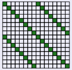
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$ 

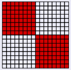
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$ 



The two-layer procedure is repeated **recursively**.

Lemma (Support of the partial products)

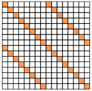
$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(4)}$

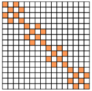
$\mathrm{supp}(\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(3)}\mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)}$

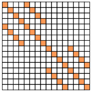
The corresponding rank-one supports are pairwise disjoint.

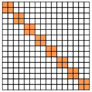
Hierarchical factorization algorithm

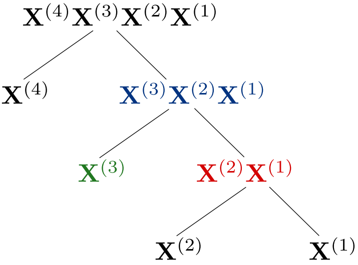
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$ 

$\mathrm{supp}(\mathbf{X}^{(3)}) \subseteq$ 

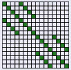
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$ 

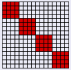
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$ 



The two-layer procedure is repeated **recursively**.

Lemma (Support of the partial products)

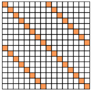
$\mathrm{supp}(\mathbf{X}^{(3)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(3)}$

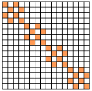
$\mathrm{supp}(\mathbf{X}^{(2)}\mathbf{X}^{(1)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(2)}\mathbf{S}_{\mathtt{bf}}^{(1)}$

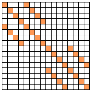
The corresponding rank-one supports are pairwise disjoint.

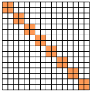
Hierarchical factorization algorithm

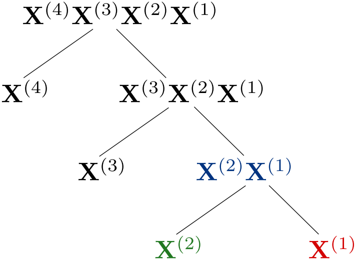
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

$\mathrm{supp}(\mathbf{X}^{(4)}) \subseteq$ 

$\mathrm{supp}(\mathbf{X}^{(3)}) \subseteq$ 

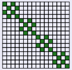
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$ 

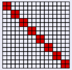
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$ 



The two-layer procedure is repeated **recursively**.

Lemma (Support of the partial products)

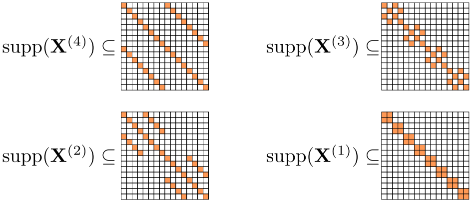
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(2)}$

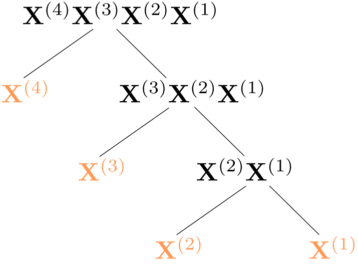
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(1)}$

The corresponding rank-one supports are pairwise disjoint.

Hierarchical factorization algorithm

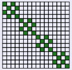
Let $\mathbf{Z} := \mathbf{X}^{(4)}\mathbf{X}^{(3)}\mathbf{X}^{(2)}\mathbf{X}^{(1)}$ such that:

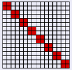




The two-layer procedure is repeated **recursively**.

Lemma (Support of the partial products)

$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(2)}$

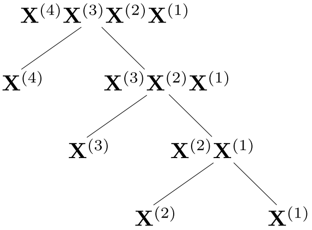
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$  $= \mathbf{S}_{\mathtt{bf}}^{(1)}$

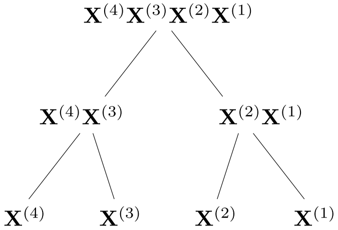
The corresponding rank-one supports are pairwise disjoint.

The butterfly factors $\{\mathbf{X}^{(\ell)}\}_{\ell=1}^{4}$ are recovered (up to scaling ambiguities) from the product $\mathbf{Z}$.

Theoretical guarantees

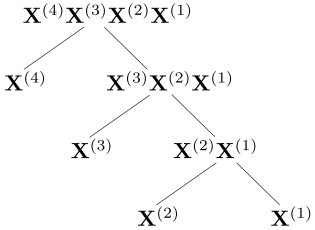
The algorithm works for any number of factors and any binary tree.





Theoretical guarantees

The algorithm works for any number of factors and any binary tree.

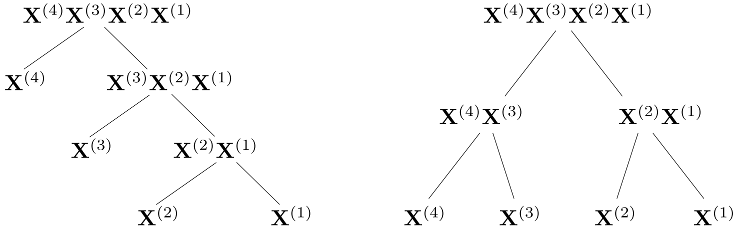


Theorem (Exact recovery guarantees [Zheng et al. 2022])

Except for trivial degeneracies, every tuple $(\mathbf{X}^{(\ell)})_{\ell=1}^{J}$ satisfying the butterfly constraint can be reconstructed by the algorithm from $\mathbf{Z} := \mathbf{X}^{(J)} \ldots \mathbf{X}^{(1)}$ (up to unavoidable scaling ambiguities).

Theoretical guarantees

The algorithm works for any number of factors and any binary tree.



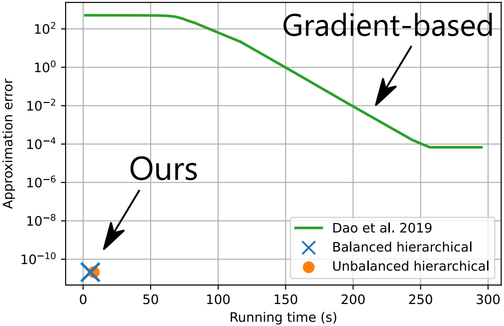
Theorem (Exact recovery guarantees [Zheng et al. 2022])

Except for trivial degeneracies, every tuple $(\mathbf{X}^{(\ell)})_{\ell=1}^{J}$ satisfying the butterfly constraint can be reconstructed by the algorithm from $\mathbf{Z} := \mathbf{X}^{(J)} \ldots \mathbf{X}^{(1)}$ (up to unavoidable scaling ambiguities).

- Complexity is $\mathcal{O}(N^2)$ for both trees.
- We can use the algorithm in the non-exact setting.

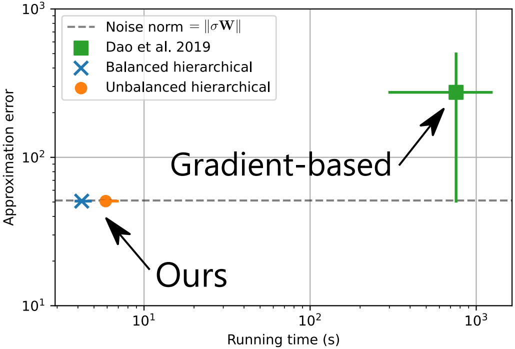
Faster and more accurate in the noiseless setting

Approximation of the DFT matrix by a product of $J = 9$ butterfly factors:



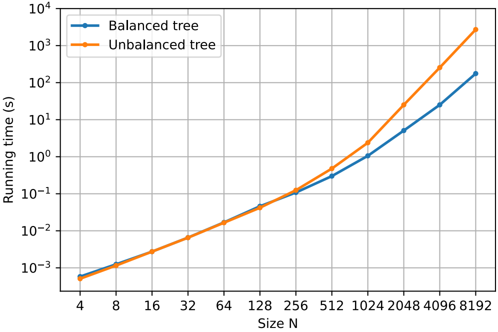
Also more robust in the noisy setting

Approximation of $\mathbf{Z} = \mathbf{DFT_N} + \sigma \mathbf{W}$ by a product of $J = 9$ butterfly factors:



Our method scales with the matrix size

Approximation of the (noisy) DFT matrix of size $N = 2^J$ by a product of J butterfly factors:



Conclusion and perspectives

Hierarchical algorithm: $\mathcal{O}(N^2)$

$\mathbf{Z} \in \mathbb{R}^{N \times N}$ (dense)

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

Cost for evaluation: $\mathcal{O}(N^2)$
$\mathbf{x} \mapsto \mathbf{Z}\mathbf{x}$

$\tilde{\mathbf{Z}} := \mathbf{X}^{(J)}\mathbf{X}^{(J-1)} \dots \mathbf{X}^{(1)}$

Storage: $\mathcal{O}(N \log N)$

Cost for evaluation: $\mathcal{O}(N \log N)$
$\mathbf{x} \mapsto \tilde{\mathbf{Z}}\mathbf{x}$

Conclusion and perspectives

Hierarchical algorithm: $\mathcal{O}(N^2)$

$\mathbf{Z} \in \mathbb{R}^{N \times N}$ (dense)

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

Cost for evaluation: $\mathcal{O}(N^2)$
$\mathbf{x} \mapsto \mathbf{Z}\mathbf{x}$

$\tilde{\mathbf{Z}} := \mathbf{X}^{(J)} \mathbf{X}^{(J-1)} \dots \mathbf{X}^{(1)}$

Storage: $\mathcal{O}(N \log N)$

Cost for evaluation: $\mathcal{O}(N \log N)$
$\mathbf{x} \mapsto \tilde{\mathbf{Z}}\mathbf{x}$

Implementation in the FAµST toolbox at `https://faust.inria.fr`.

Conclusion and perspectives

Hierarchical algorithm: $\mathcal{O}(N^2)$

$\mathbf{Z} \in \mathbb{R}^{N \times N}$ (dense)

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

Cost for evaluation: $\mathcal{O}(N^2)$
$\mathbf{x} \mapsto \mathbf{Z}\mathbf{x}$

$\tilde{\mathbf{Z}} := \mathbf{X}^{(J)}\mathbf{X}^{(J-1)} \dots \mathbf{X}^{(1)}$

Storage: $\mathcal{O}(N \log N)$

Cost for evaluation: $\mathcal{O}(N \log N)$
$\mathbf{x} \mapsto \tilde{\mathbf{Z}}\mathbf{x}$

Implementation in the FAµST toolbox at `https://faust.inria.fr`.

Future work

- Application in dictionary learning, sparse neural network training, ...
- Stability properties of the hierarchical algorithm

Thank you for your attention!

To know more:

Q.-T. Le, L. Zheng, E. Riccietti, and R. Gribonval (2022)
Fast learning of fast transforms, with guarantees
In ICASSP, 2022.

Q.-T. Le, E. Riccietti, and R. Gribonval (2022)
Spurious Valleys, Spurious Minima and NP-hardness of Sparse Matrix Factorization With Fixed Support
arXiv preprint, arXiv:2112.00386.

L. Zheng, E. Riccietti, and R. Gribonval (2022)
Efficient Identification of Butterfly Sparse Matrix Factorizations
arXiv preprint, arXiv:2110.01235.