Fixed support matrix factorization

Quoc Tung Le   Léon Zheng   Elisa Riccietti   Rémi Gribonval

Université de Lyon, Ecole Normale Supérieure de Lyon

INRIA, CNRS, Laboratoire de l'Informatique du Parallélisme - LIP

Sparse Days, St. Girons, June 20-22, 2022

Sparse matrix factorization

Given a dense matrix A, find *multiple* factors $S_1, S_2, \ldots S_J$ such that:

$$A \approx S_1 S_2 \ldots S_J$$

where S_i are *sparse* matrices.

Sparse matrix factorization

Given a dense matrix A, find *multiple* factors $S_1, S_2, \dots S_J$ such that:

$$A \approx S_1 S_2 \dots S_J$$

where S_i are *sparse* matrices.

$$\underbrace{A}_{\textit{dense}} \approx \underbrace{S_1 S_2 \dots S_J}_{\textit{sparse}} \quad \Rightarrow \quad Ax \approx S_1(S_2(\dots(S_J x)))$$

Sparse matrix factorization

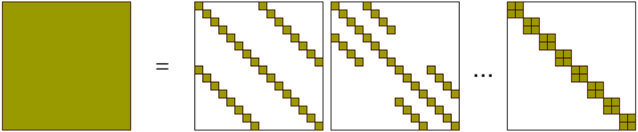
Given a dense matrix A, find *multiple* factors $S_1, S_2, \dots S_J$ such that:

$$A \approx S_1 S_2 \dots S_J$$

where S_i are *sparse* matrices.

$$\underbrace{A}_{dense} \approx \underbrace{S_1 S_2 \dots S_J}_{sparse} \quad \Rightarrow \quad Ax \approx S_1(S_2(\dots(S_J x)))$$

Application: the Fast Fourier Transform, the Fast Hadamard Transform, etc.



The factorization of the Discrete Fourier Transform.

A general formulation for sparse matrix factorization

Sparse Matrix Factorization Problem

Given a matrix A, $J \in \mathbb{N}$ and $\mathcal{E}_j$ some sets of sparse matrices, solve:

$$\begin{aligned} \underset{S_1,\ldots,S_J}{\text{Minimize}} \quad & \|A - \prod_{j=1}^{J} S_j\|_F^2 \\ \text{subject to:} \quad & S_j \in \mathcal{E}_j, \qquad \forall j \in \{1,\ldots,J\} \end{aligned}$$

A general formulation for sparse matrix factorization

Sparse Matrix Factorization Problem

Given a matrix A, $J \in \mathbb{N}$ and $\mathcal{E}_j$ some sets of sparse matrices, solve:

$$\underset{S_1,\ldots,S_J}{\text{Minimize}} \quad \|A - \prod_{j=1}^{J} S_j\|_F^2$$

$$\text{subject to:} \quad S_j \in \mathcal{E}_j, \qquad \forall j \in \{1,\ldots,J\}$$

Examples of matrix sets $\mathcal{E}$:

1) $\mathcal{E}^k_{row} = \{S : |\mathrm{supp}(S_{i,\bullet})| \leq k\}$: at most k nonzero entries per row.
2) $\mathcal{E}^k_{col} = \{S : |\mathrm{supp}(S_{\bullet,i})| \leq k\}$: at most k nonzero entries per column.
3) $\mathcal{E}^k_{tot} = \{S : |\mathrm{supp}(S)| \leq k\}$: at most k nonzero entries in total.

A general formulation for sparse matrix factorization

Sparse Matrix Factorization Problem

Given a matrix A, $J \in \mathbb{N}$ and $\mathcal{E}_j$ some sets of sparse matrices, solve:

$$\underset{S_1,\ldots,S_J}{\text{Minimize}} \quad \|A - \prod_{j=1}^{J} S_j\|_F^2$$

$$\text{subject to:} \quad S_j \in \mathcal{E}_j, \qquad \forall j \in \{1, \ldots, J\}$$

Examples of matrix sets $\mathcal{E}$:

1) $\mathcal{E}_{row}^k = \{S : |\mathrm{supp}(S_{i,\bullet})| \leq k\}$: at most k nonzero entries per row.
2) $\mathcal{E}_{col}^k = \{S : |\mathrm{supp}(S_{\bullet,i})| \leq k\}$: at most k nonzero entries per column.
3) $\mathcal{E}_{tot}^k = \{S : |\mathrm{supp}(S)| \leq k\}$: at most k nonzero entries in total.

$\rightarrow$ A challenging problem, how to deal with it?

A classical related problem: sparse linear inverse problem

$$\underset{y \in \mathbb{R}^n}{\text{Minimize}} \quad \|a - Xy\|_2^2 \qquad \text{subject to:} \qquad \|y\|_0 \leq s, s \ll n$$

A classical related problem: sparse linear inverse problem

$$\underset{y \in \mathbb{R}^n}{\text{Minimize}} \quad \|a - Xy\|_2^2 \qquad \text{subject to:} \qquad \|y\|_0 \leq s, s \ll n$$

1) Support identification

Finding a set $I \subseteq [\![n]\!]$ such that $|I| = s$.

2) Optimize coefficients inside support

$$\underset{y \in \mathbb{R}^n, \mathrm{supp}(y) \subseteq I}{\text{Minimize}} \|a - Xy\|_2^2$$

A classical related problem: sparse linear inverse problem

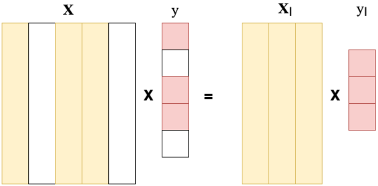
$$\underset{y \in \mathbb{R}^n}{\text{Minimize}} \quad \|a - Xy\|_2^2 \qquad \text{subject to:} \qquad \|y\|_0 \leq s, s \ll n$$

1) Support identification

Finding a set $I \subseteq [\![n]\!]$ such that $|I| = s$.

2) Optimize coefficients inside support

$$\underset{y \in \mathbb{R}^n, \mathrm{supp}(y) \subseteq I}{\text{Minimize}} \|a - Xy\|_2^2 = \|a - X_I y_I\|_2^2$$



A classical related problem: sparse linear inverse problem

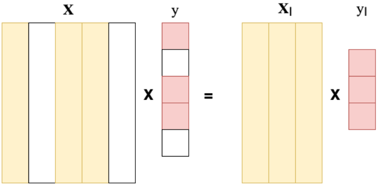
$$\underset{y \in \mathbb{R}^n}{\text{Minimize}} \quad \|a - Xy\|_2^2 \qquad \text{subject to:} \qquad \|y\|_0 \leq s, s \ll n$$

1) Support identification

Finding a set $I \subseteq [\![n]\!]$ such that $|I| = s$.

2) *Linear regression problem*

$$\underset{y_I \in \mathbb{R}^{|I|}}{\text{Minimize}} \quad \|a - X_I y_I\|_2^2$$



Two sub-problems of **two** factors matrix factorization

$$\underset{X,Y}{\text{Minimize}} \quad \|A - XY^\top\|_F^2 \quad \text{subject to: } X, Y \text{ sparse matrices}$$

$\rightarrow$ The simplest non-trivial setting.

Two sub-problems of **two** factors matrix factorization

$$\underset{X,Y}{\text{Minimize}} \quad \|A - XY^\top\|_F^2 \quad \text{subject to: } X, Y \text{ sparse matrices}$$

→ The simplest non-trivial setting.

1) Support identification

Find *two* sets $I \subseteq [\![m]\!] \times [\![r]\!]$ and $J \subseteq [\![n]\!] \times [\![r]\!]$ satisfying the sparse matrix sets constraint $\mathcal{E}$ such that $\operatorname{supp}(X) \subseteq I, \operatorname{supp}(Y) \subseteq J$.

Two sub-problems of **two** factors matrix factorization

$$\underset{X,Y}{\text{Minimize}} \quad \|A - XY^\top\|_F^2 \quad \text{subject to: } X, Y \text{ sparse matrices}$$

$\rightarrow$ The simplest non-trivial setting.

1) Support identification

Find *two* sets $I \subseteq [\![m]\!] \times [\![r]\!]$ and $J \subseteq [\![n]\!] \times [\![r]\!]$ satisfying the sparse matrix sets constraint $\mathcal{E}$ such that $\mathrm{supp}(X) \subseteq I, \mathrm{supp}(Y) \subseteq J$.

2) Optimize coefficients inside support

$$\begin{aligned} \underset{X \in \mathbb{R}^{m \times r}, Y \in \mathbb{R}^{n \times r}}{\text{Minimize}} \quad & L(X, Y) = \|A - XY^\top\|^2 \\ \text{Subject to:} \quad & \mathrm{supp}(X) \subseteq I \\ & \mathrm{supp}(Y) \subseteq J \end{aligned}$$

A comparison between two problems

	Linear inverse problem	*Sparse matrix factorization*
Pb	Minimize $\|a - Xy\|$, a, X are *known*, y is sparse	Minimize $\|A - XY\|$, A is *known*, X, Y are sparse
1)	*Hard* due to exponential growth of combinations	
2)	*Easy* - Linear regression problem	*Unknown* to the literature

A comparison between two problems

	Linear inverse problem	*Sparse matrix factorization*
Pb	Minimize $\|a - Xy\|$, a, X are *known*, y is sparse	Minimize $\|A - XY\|$, A is *known*, X, Y are sparse
1)	*Hard* due to exponential growth of combinations	
2)	*Easy* - Linear regression problem	FSMF

Fixed support matrix factorization

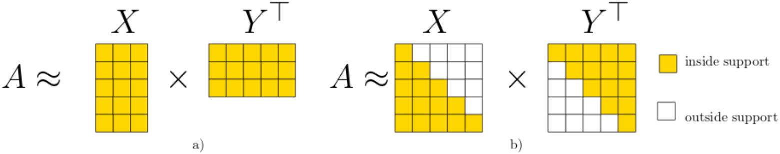
$$\underset{X \in \mathbb{R}^{m \times r}, Y \in \mathbb{R}^{n \times r}}{\text{Minimize}} \quad L(X, Y) = \|A - XY^\top\|^2$$

$$\text{Subject to:} \quad \text{supp}(X) \subseteq I, \quad \text{supp}(Y) \subseteq J \qquad \text{(FSMF)}$$

FSMF: motivation (I)

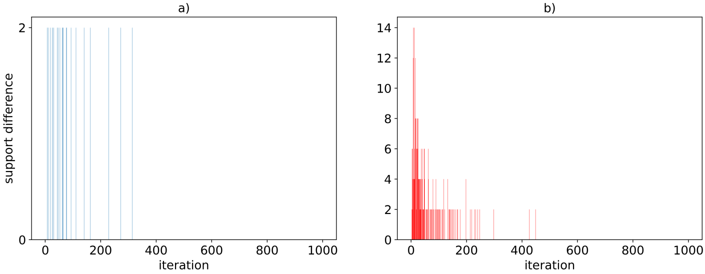
Fixed support matrix factorization covers several existing frameworks:

- Low rank matrix decomposition
- LU decomposition
- Hierarchical $\mathcal{H}$ and BLR matrices
- Butterfly factorization



FSMF: motivation (II)

FSMF helps to understand the *asymptotic behaviour* of heuristics such as PALM: alternate update of the factors by projected gradient step onto the set of the constraints.



Main contributions

Fixed support matrix factorization

$$\begin{aligned} \underset{X\in\mathbb{R}^{m\times r}, Y\in\mathbb{R}^{n\times r}}{\text{Minimize}} \quad & L(X,Y) = \|A - XY^\top\|^2 \\ \text{Subject to:} \quad & \operatorname{supp}(X) \subseteq I \\ & \operatorname{supp}(Y) \subseteq J \end{aligned} \tag{FSMF}$$

1) Is the problem in P (polynomially tractable)?
   $\rightarrow$ We have proved its *NP-hardness*.
2) Are there easy instances?
   $\rightarrow$ We individuated a family of *polynomially solvable* instances, proved the well-posedness of the problem and proposed an efficient algorithm.
3) How well does gradient descent tackle the problem of FSMF?
   $\rightarrow$ We have studied the properties of the *landscape* of the function $L(X,Y) = \|A - XY^\top\|^2$ *under the support constraints*.

Main contributions

Fixed support matrix factorization

$$
\begin{aligned}
\underset{X\in\mathbb{R}^{m\times r}, Y\in\mathbb{R}^{n\times r}}{\text{Minimize}} \quad & L(X,Y) = \|A - XY^\top\|^2 \\
\text{Subject to:} \quad & \mathrm{supp}(X) \subseteq I \\
& \mathrm{supp}(Y) \subseteq J
\end{aligned}
\tag{FSMF}
$$

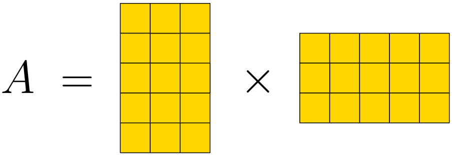
2) Otherwise, are there easy instances?
→ We individuated several *polynomially solvable* cases and proposed an efficient algorithm.

Polynomially solvable instances

Example (Unconstrained matrix factorization)

If $I = [\![m]\!] \times [\![r]\!]$, $J = [\![n]\!] \times [\![r]\!]$, i.e no constraints on the support of X and Y:

$$\underset{X \in \mathbb{R}^{m \times r}, Y \in \mathbb{R}^{n \times r}}{\text{Minimize}} L(X, Y) = \|A - XY^\top\|^2$$

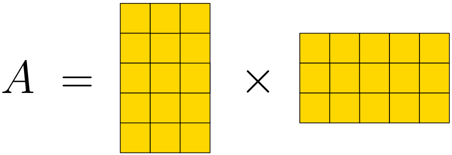


Polynomially solvable instances

Example (Unconstrained matrix factorization)

If $I = [\![m]\!] \times [\![r]\!]$, $J = [\![n]\!] \times [\![r]\!]$, i.e no constraints on the support of X and Y:

$$\underset{X \in \mathbb{R}^{m \times r}, Y \in \mathbb{R}^{n \times r}}{\text{Minimize}} L(X, Y) = \|A - XY^\top\|^2$$



$\rightarrow$ Solution: Use Singular Value Decomposition (SVD).

SVD as a greedy algorithm

1) Decompose the problem:

$$A - XY^\top = A - \sum_{i=1}^{r} x_i y_i^\top = A - \sum_{i=1}^{r} \underbrace{M_i}_{\text{rank one}} \quad (M_i := x_i y_i^\top)$$

SVD as a greedy algorithm

1) Decompose the problem:

$$A - XY^\top = A - \sum_{i=1}^{r} x_i y_i^\top = A - \sum_{i=1}^{r} \underbrace{M_i}_{\text{rank one}} \quad (M_i := x_i y_i^\top)$$

2) Finding the SVD:

$$\begin{aligned}
&\texttt{bestRankOneApprox}(A) && \to M_1 \\
&\texttt{bestRankOneApprox}(A - M_1) && \to M_2 \\
&\cdots && \\
&\texttt{bestRankOneApprox}(A - M_1 \ldots - M_{r-1}) && \to M_r
\end{aligned}$$

SVD as a greedy algorithm

1) Decompose the problem:

$$A - XY^\top = A - \sum_{i=1}^{r} x_i y_i^\top = A - \sum_{i=1}^{r} \underbrace{M_i}_{\text{rank one}} \qquad (M_i := x_i y_i^\top)$$

2) Finding the SVD:

$$\begin{aligned}
&\texttt{bestRankOneApprox}(A) && \rightarrow M_1 \\
&\texttt{bestRankOneApprox}(A - M_1) && \rightarrow M_2 \\
&\dots && \\
&\texttt{bestRankOneApprox}(A - M_1 \dots - M_{r-1}) && \rightarrow M_r
\end{aligned}$$

$\rightarrow$ SVD is a greedy algorithm in disguise

Algorithm 1 Algorithm for unconstrained matrix factorization

---

for $i \in \{1, \dots, r\}$ **do**
    $M_i :=$ best rank one approximation of $A - \sum_{k=1}^{i-1} M_k$.
end for

Notions of rank one support

How to generalize the greedy algorithm?

Notions of rank one support

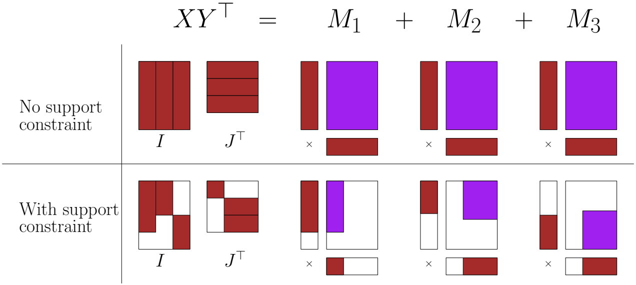
How to generalize the greedy algorithm?
Decompose $XY^\top$:

$$XY^\top = \sum_{i=1}^{r} X_i Y_i^\top = \sum_{i=1}^{r} \underbrace{M_i}_{\text{rank one}} \quad (M_i := X_i Y_i^\top)$$

Notions of rank one support

How to generalize the greedy algorithm?
Decompose $XY^\top$:

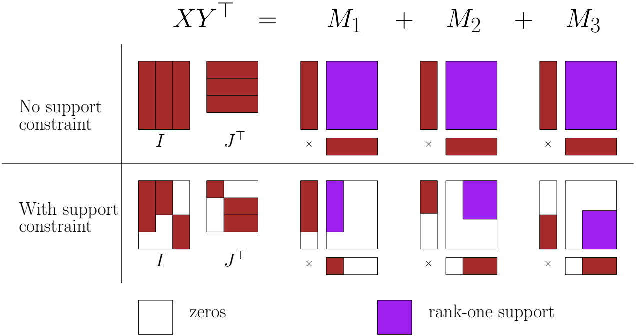
$$XY^\top = \sum_{i=1}^{r} X_i Y_i^\top = \sum_{i=1}^{r} \underbrace{M_i}_{\text{rank one}} \qquad (M_i := X_i Y_i^\top)$$



Notions of rank one support

How to generalize the greedy algorithm?
Decompose $XY^\top$:

$$XY^\top = \sum_{i=1}^{r} X_i Y_i^\top = \sum_{i=1}^{r} \underbrace{M_i}_{\text{rank one}} \qquad (M_i := X_i Y_i^\top)$$

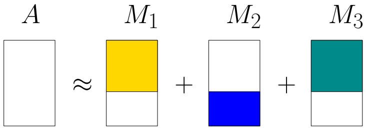


Finding optimal solution (X, Y) $\rightleftarrows$ Finding optimal entries in rank-one support.

Polynomial solvability characterized by rank-one supports

Theorem (Sufficient condition for tractability (Informal))

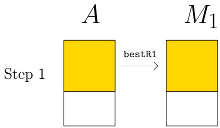
If the rank-one supports are pairwise disjoint or identical, the problem admits an essentially unique solution and the greedy algorithm gives the optimal solution.



Polynomial solvability characterized by rank-one supports

Theorem (Sufficient condition for tractability (Informal))

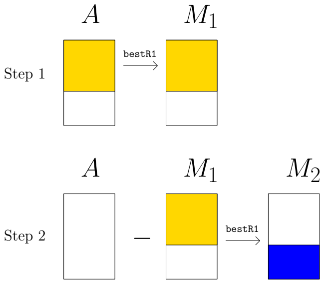
If the rank-one supports are pairwise disjoint or identical, the problem admits an essentially unique solution and the greedy algorithm gives the optimal solution.



Polynomial solvability characterized by rank-one supports

Theorem (Sufficient condition for tractability (Informal))

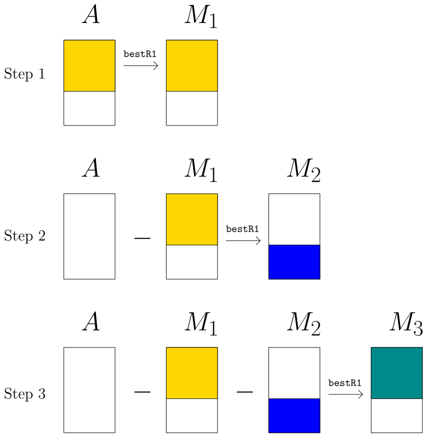
If the rank-one supports are pairwise disjoint or identical, the problem admits an essentially unique solution and the greedy algorithm gives the optimal solution.



Polynomial solvability characterized by rank-one supports

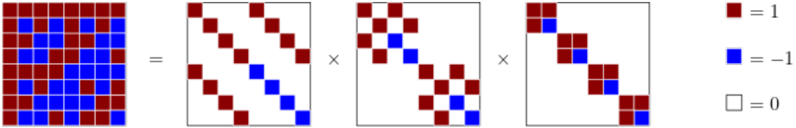
Theorem (Sufficient condition for tractability (Informal))

If the rank-one supports are pairwise disjoint or identical, the problem admits an essentially unique solution and the greedy algorithm gives the optimal solution.

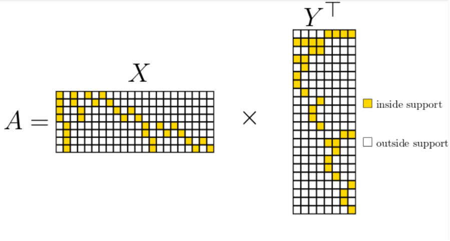


Examples

- Butterfly supports: Discrete Fourier Transform (DFT) or the Hadamard transform (HT)



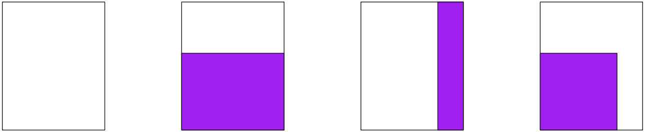
- Hierarchically off-diagonal low-rank (HODLR) matrices



An even more general result exists

- A more general condition for tractability is introduced in our paper that allows partial overlapping[1]

$$A \approx M_1 + M_2 + M_3$$

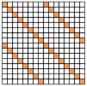


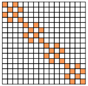
[1] Quoc-Tung Le, Elisa Riccietti, and Rémi Gribonval. "Spurious Valleys, Spurious Minima and NP-hardness of Sparse Matrix Factorization With Fixed Support". working paper or preprint. May 2021. URL: https://hal.inria.fr/hal-03364668.

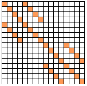
Hierarchical factorization algorithm

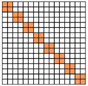
Extension: an efficient **hierarchical algorithm** to approximate **any** matrix by a product of $J \geq 2$ butterfly factors.

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

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

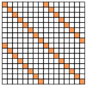
$\text{supp}(\mathbf{X}^{(3)}) \subseteq$ 

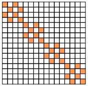
$\text{supp}(\mathbf{X}^{(2)}) \subseteq$ 

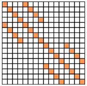
$\text{supp}(\mathbf{X}^{(1)}) \subseteq$ 

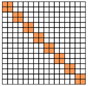
Extension: an efficient **hierarchical algorithm** to approximate **any** matrix by a product of $J \geq 2$ butterfly factors.

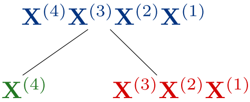
Let $\mathrm{A} := \mathrm{X}^{(4)}\mathrm{X}^{(3)}\mathrm{X}^{(2)}\mathrm{X}^{(1)}$ such that:

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

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

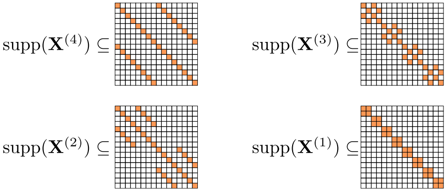
$\mathrm{supp}(\mathbf{X}^{(2)}) \subseteq$ 

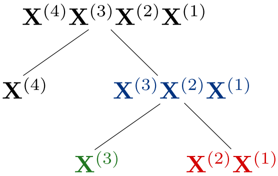
$\mathrm{supp}(\mathbf{X}^{(1)}) \subseteq$ 



Extension: an efficient **hierarchical algorithm** to approximate **any** matrix by a product of $J \geq 2$ butterfly factors.

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

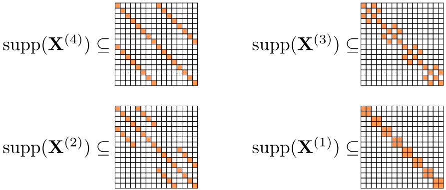


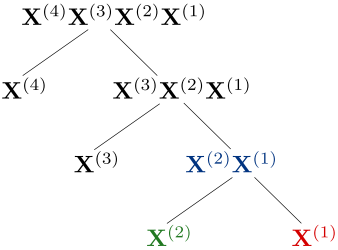


Hierarchical factorization algorithm

Extension: an efficient **hierarchical algorithm** to approximate **any** matrix by a product of $J \geq 2$ butterfly factors.

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

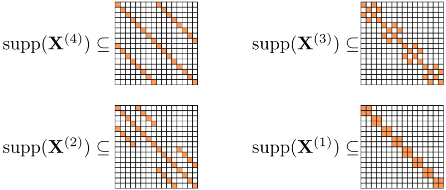


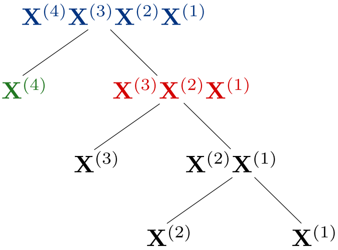


Hierarchical factorization algorithm

Extension: an efficient **hierarchical algorithm** to approximate **any** matrix by a product of $J \geq 2$ butterfly factors.

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



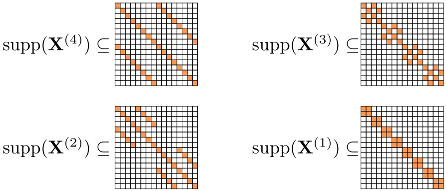


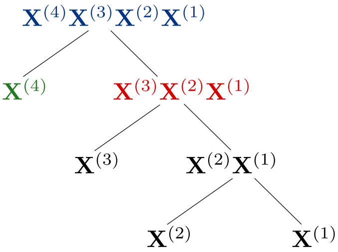
How to recover the partial products?

Hierarchical factorization algorithm

Extension: an efficient **hierarchical algorithm** to approximate **any** matrix by a product of $J \geq 2$ butterfly factors.

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





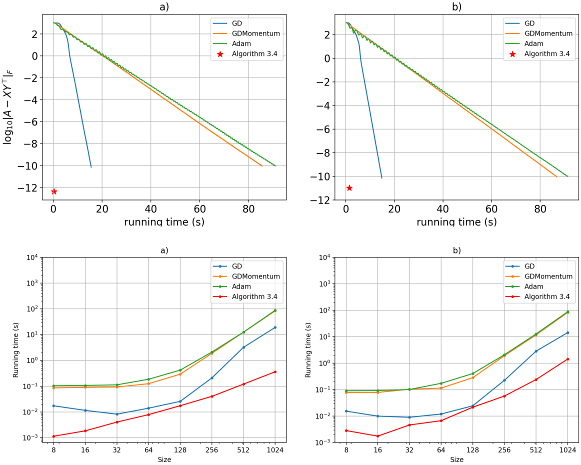
How to recover the partial products? $\rightarrow$ use their known supports

Lemma (Supports of the partial products)

At each level the rank-one matrices have pairwise disjoint supports. We can use our algorithm to recover the partial factors: solve a sequence of two factors problems

Numerical results: 2 factors

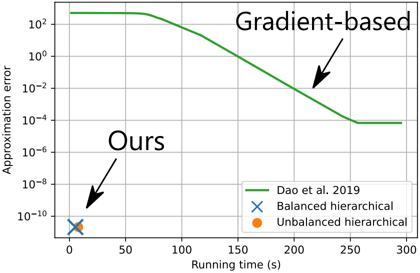
A the Hadamard matrix of size $2^N \times 2^N$, $a = \lceil N/2 \rceil, b = \lfloor N/2 \rfloor$, $N = 10$



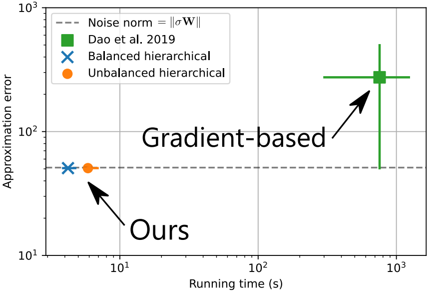
Numerical results: J factors

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

Faster and more accurate in the noiseless setting



Also more robust in the noisy setting



Conclusion

Take home message

For Fixed support matrix factorization, we have:

1) It is NP-hard to solve.
2) Easy instances with effective direct algorithm exists, competitive with gradient descent

Conclusion

Take home message

For Fixed support matrix factorization, we have:

1) It is NP-hard to solve.
2) Easy instances with effective direct algorithm exists, competitive with gradient descent

Perspectives

Theory and algorithms for multiple factors[a].

[a] Quoc-Tung Le et al. "Fast learning of fast transforms, with guarantees". In: *ICASSP 2022 - IEEE International Conference on Acoustics, Speech and Signal Processing*. Singapore, Singapore, May 2022.

To know more:

- 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.

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