Analyzing identifiability of sparse linear networks

GdR ISIS - Theory of deep learning

Léon Zheng, Rémi Gribonval, Elisa Riccietti

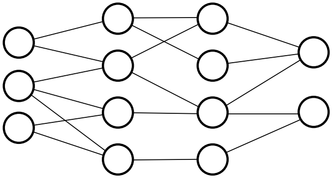
Inria DANTE / ENS de Lyon (LIP)

June 28, 2021

Sparse (linear) neural networks

- Reduce time + space complexity
- Toward interpretable NN?

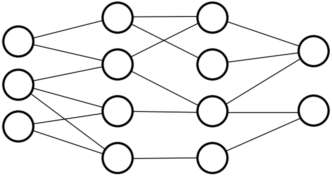
→ requires identifiability / stability



Sparse (linear) neural networks

- Reduce time + space complexity
- Toward interpretable NN?

→ requires identifiability / stability



Analogy with NMF

Identifiability ensures that solution to NMF can be interpreted as the physical ground-truth.

Example: blind hyperspectral unmixing.

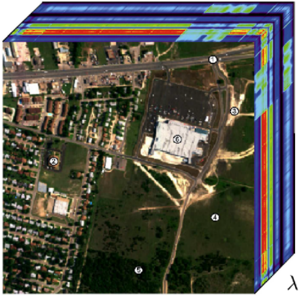


Figure: from [Gillis 2020]

Well-posedness in sparse matrix factorization?

Given a matrix $\boldsymbol{Z}$, and $L \geq 2$, solve

$$\min_{\boldsymbol{X_1},\dots,\boldsymbol{X_L}} \|\boldsymbol{Z} - \boldsymbol{X_L}\boldsymbol{X_{L-1}}\dots\boldsymbol{X_1}\|$$
$$\text{such that} \quad \boldsymbol{X_\ell} \text{ is } \textit{sparse}, \quad \forall \ell \in \{1, \dots, L\},$$

by exploring a given family of supports, with proximal algorithm [Le Magoarou and Gribonval 2016].

Well-posedness in sparse matrix factorization?

Given a matrix $\boldsymbol{Z}$, and $L \geq 2$, solve

$$\min_{\boldsymbol{X_1},\ldots,\boldsymbol{X_L}} \|\boldsymbol{Z} - \boldsymbol{X_L}\boldsymbol{X_{L-1}}\ldots\boldsymbol{X_1}\|$$

$$\text{such that} \quad \boldsymbol{X_\ell} \text{ is } \textit{sparse}, \quad \forall \ell \in \{1, \ldots, L\},$$

by exploring a given family of supports, with proximal algorithm [Le Magoarou and Gribonval 2016].

Condition of success?

Well-posedness of the problem is the key to recovery success:

- **uniqueness** of the solution to recover
- **stability** with respect to noise

Well-posedness in sparse matrix factorization?

Given a matrix $\boldsymbol{Z}$, and $L \geq 2$, solve

$$\min_{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_L} \|\boldsymbol{Z} - \boldsymbol{X}_L \boldsymbol{X}_{L-1} \ldots \boldsymbol{X}_1\|$$

$$\text{such that} \quad \boldsymbol{X}_\ell \text{ is } \textit{sparse}, \quad \forall \ell \in \{1, \ldots, L\},$$

by exploring a given family of supports, with proximal algorithm [Le Magoarou and Gribonval 2016].

Condition of success?

Well-posedness of the problem is the key to recovery success:

- **uniqueness** of the solution to recover
- **stability** with respect to noise

Focus on uniqueness in **exact** sparse matrix factorization → identifiability

Identifiability in exact sparse matrix facotrization

Outline

1. Analysis with two factors
2. Multilayer case via hierarchical factorization method

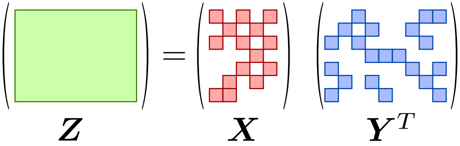
Identifiability in exact sparse matrix facotrization

Outline

1. Analysis with two factors
2. Multilayer case via hierarchical factorization method

Given a matrix $\boldsymbol{Z}$ and a feasible set $\Sigma^L \times \Sigma^R$ of pairs of factors, define:

$$\text{find, if possible, } (\boldsymbol{X}, \boldsymbol{Y}) \in \Sigma^L \times \Sigma^R \text{ such that } \boldsymbol{Z} = \boldsymbol{X}\boldsymbol{Y}^T. \qquad \text{(EMF)}$$



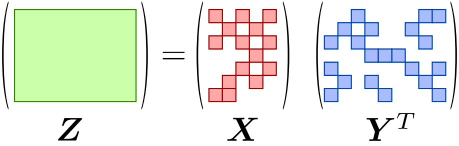
Identifiability in exact sparse matrix facotrization

Outline

1. Analysis with two factors
2. Multilayer case via hierarchical factorization method

Given a matrix $\boldsymbol{Z}$ and a feasible set $\Sigma^L \times \Sigma^R$ of pairs of factors, define:

$$\text{find, if possible, } (\boldsymbol{X}, \boldsymbol{Y}) \in \Sigma^L \times \Sigma^R \text{ such that } \boldsymbol{Z} = \boldsymbol{X}\boldsymbol{Y}^T. \qquad \text{(EMF)}$$



Informal Theorem

Let $\boldsymbol{Z}$ be a matrix, and $\Sigma^L \times \Sigma^R$ encoding sparsity on pairs of factors. If a certain condition on $\Sigma^L \times \Sigma^R$, then $\boldsymbol{Z}$ admits a unique EMF $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{Y}^T$ in $\Sigma^L \times \Sigma^R$, up to scaling and permutation ambiguities.

Identifiability in exact sparse matrix facotrization

Outline

1. Analysis with two factors
2. Multilayer case via hierarchical factorization method

Given a matrix $\boldsymbol{Z}$ and a feasible set $\Sigma^L \times \Sigma^R$ of pairs of factors, define:

$$\text{find, if possible, } (\boldsymbol{X}, \boldsymbol{Y}) \in \Sigma^L \times \Sigma^R \text{ such that } \boldsymbol{Z} = \boldsymbol{X}\boldsymbol{Y}^T. \qquad \text{(EMF)}$$

Theorem

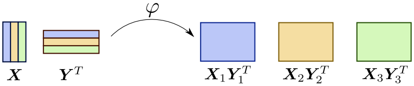
Let $\boldsymbol{Z}$ be the DFT, DCT-II or DST-II matrix of size $N = 2^L$. Suppose that:

- *Σ^L enforces 2-sparsity by column;*
- *Σ^R enforces $\frac{N}{2}$-sparsity by column.*

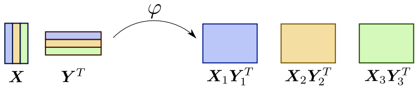
Then, $\boldsymbol{Z}$ admits a unique EMF $\boldsymbol{Z} = \boldsymbol{X}\boldsymbol{Y}^T$ in $\Sigma^L \times \Sigma^R$, up to scaling and permutation ambiguities.

Notation: $\Sigma^2_{\texttt{col}} \times \Sigma^{N/2}_{\texttt{col}}$.

Matrix decomposition into sparse rank-one matrices



Matrix decomposition into sparse rank-one matrices

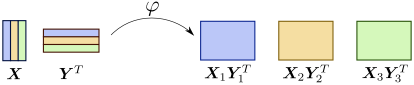


Given a matrix $\mathbf{Z}$ and a feasible set Γ of r-tuples of rank-one matrices, define:

$$\text{find, if possible, } (\mathcal{C}^i)_{i=1}^r \in \Gamma \text{ such that } \mathbf{Z} = \sum_{i=1}^{r} \mathcal{C}^i. \qquad \text{(EMD)}$$

$\rightarrow$ lifting procedure [Choudhary and Mitra 2014], [Le Magoarou 2016]

Matrix decomposition into sparse rank-one matrices



Given a matrix $\boldsymbol{Z}$ and a feasible set Γ of r-tuples of rank-one matrices, define:

$$\text{find, if possible, } (\mathcal{C}^i)_{i=1}^r \in \Gamma \text{ such that } \boldsymbol{Z} = \sum_{i=1}^{r} \mathcal{C}^i. \tag{EMD}$$

$\rightarrow$ lifting procedure [Choudhary and Mitra 2014], [Le Magoarou 2016]

Proposition

When $(\boldsymbol{X}, \boldsymbol{Y})$ is non-degenerate, identifiability of $(\boldsymbol{X}, \boldsymbol{Y})$ for the EMF of $\boldsymbol{Z} := \boldsymbol{X}\boldsymbol{Y}^T$ in $\Sigma^L \times \Sigma^R$ is equivalent to identifiability of $\varphi(\boldsymbol{X}, \boldsymbol{Y})$ for the EMD of $\boldsymbol{Z}$ in Γ.

In the case of $\Sigma^2_{\texttt{col}} \times \Sigma^{N/2}_{\texttt{col}}$:

$$\Gamma^{2,N/2} := \left\{ (\mathcal{C}^i)_{i=1}^r \mid \mathcal{C}^i \text{ has 2 nonzero rows, } \frac{N}{2} \text{ nonzero columns} \right\}.$$

Fixed-support identifiability

Analogy with sparse linear recovery (recover s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$):

- identifiability of the support constraint
- fixed-support identifiability

Fixed-support identifiability

Analogy with sparse linear recovery (recover s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$):

- identifiability of the support constraint
- fixed-support identifiability

Proposition

When the rank-one supports "do not overlay too much", it is possible to complete without ambiguity missing entries from observable entries via rank-one matrix completion.

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \\ 2 & 6 & 5 & 6 \\ 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ ? & ? & 0 & 0 \\ ? & ? & 0 & 0 \\ ? & ? & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & ? & ? & 0 \\ 0 & ? & ? & 0 \\ 0 & ? & ? & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & ? & ? & ? \\ 0 & ? & ? & ? \end{pmatrix}$$

Fixed-support identifiability

Analogy with sparse linear recovery (recover s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$):

- identifiability of the support constraint
- fixed-support identifiability

Proposition

When the rank-one supports "do not overlay too much", it is possible to complete without ambiguity missing entries from observable entries via rank-one matrix completion.

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \\ 2 & 6 & 5 & 6 \\ 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & ? & 0 & 0 \\ 2 & ? & 0 & 0 \\ 3 & ? & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & ? & 2 & 0 \\ 0 & ? & ? & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & ? & ? & 6 \\ 0 & ? & 2 & 4 \end{pmatrix}$$

Fixed-support identifiability

Analogy with sparse linear recovery (recover s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$):

- identifiability of the support constraint
- fixed-support identifiability

Proposition

When the rank-one supports "do not overlay too much", it is possible to complete without ambiguity missing entries from observable entries via rank-one matrix completion.

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \\ 2 & 6 & 5 & 6 \\ 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & ? & 0 & 0 \\ 2 & ? & 0 & 0 \\ 3 & ? & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & ? & ? & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & ? & 3 & 6 \\ 0 & ? & 2 & 4 \end{pmatrix}$$

Fixed-support identifiability

Analogy with sparse linear recovery (recover s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$):

- identifiability of the support constraint
- fixed-support identifiability

Proposition

When the rank-one supports "do not overlay too much", it is possible to complete without ambiguity missing entries from observable entries via rank-one matrix completion.

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \\ 2 & 6 & 5 & 6 \\ 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & ? & 0 & 0 \\ 3 & ? & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & ? & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & ? & 3 & 6 \\ 0 & ? & 2 & 4 \end{pmatrix}$$

Fixed-support identifiability

Analogy with sparse linear recovery (recover s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$):

- identifiability of the support constraint
- fixed-support identifiability

Proposition

When the rank-one supports "do not overlay too much", it is possible to complete without ambiguity missing entries from observable entries via rank-one matrix completion.

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \\ 2 & 6 & 5 & 6 \\ 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 3 & 3 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & ? & 3 & 6 \\ 0 & ? & 2 & 4 \end{pmatrix}$$

Fixed-support identifiability

Analogy with sparse linear recovery (recover s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$):

- identifiability of the support constraint
- fixed-support identifiability

Proposition

When the rank-one supports "do not overlay too much", it is possible to complete without ambiguity missing entries from observable entries via rank-one matrix completion.

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \\ 2 & 6 & 5 & 6 \\ 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 3 & 3 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 3 & 3 & 6 \\ 0 & 2 & 2 & 4 \end{pmatrix}$$

Fixed-support identifiability

Analogy with sparse linear recovery (recover s-sparse $\boldsymbol{x}$ from $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$):

- identifiability of the support constraint
- fixed-support identifiability

Proposition

When the rank-one supports "do not overlay too much", it is possible to complete without ambiguity missing entries from observable entries via rank-one matrix completion.

$$\begin{pmatrix} 0 & 1 & 2 & 0 \\ 1 & 2 & 2 & 0 \\ 2 & 6 & 5 & 6 \\ 3 & 5 & 2 & 4 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 2 & 2 & 0 & 0 \\ 3 & 3 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 3 & 3 & 6 \\ 0 & 2 & 2 & 4 \end{pmatrix}$$

Remark: condition verified when the rank-one supports are disjoint.

Identifying the support constraint

Proposition

Let $\boldsymbol{Z}$ be the DFT, DCT-II or DST-II matrix. Then, for any EMD $\boldsymbol{Z} = \sum_{i=1}^{r} \boldsymbol{\mathcal{C}}^i$ with $\mathcal{C} \in \Gamma^{2,N/2}$, there exists σ such that: $\mathrm{supp}(\boldsymbol{\mathcal{C}}^i) \subseteq \boldsymbol{\mathcal{S}}^{\sigma(i)}$, where $\{\boldsymbol{\mathcal{S}}^i\}_{i=1}^r$ are pairwise disjoint.

$$\boldsymbol{DFT_4} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$$

Identifying the support constraint

Proposition

Let $\boldsymbol{Z}$ be the DFT, DCT-II or DST-II matrix. Then, for any EMD $\boldsymbol{Z} = \sum_{i=1}^{r} \boldsymbol{\mathcal{C}}^i$ with $\mathcal{C} \in \Gamma^{2,N/2}$, there exists σ such that: $\text{supp}(\boldsymbol{\mathcal{C}}^i) \subseteq \boldsymbol{\mathcal{S}}^{\sigma(i)}$, *where $\{\boldsymbol{\mathcal{S}}^i\}_{i=1}^r$ are pairwise disjoint.*

$$\boldsymbol{DFT_4} = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$$

1. If $\mathcal{C} \in \Gamma^{2,N/2}$ is an EMD of $\boldsymbol{DFT_4}$, then $16 = \|\boldsymbol{DFT_4}\|_0 = \|\sum_{i=1}^4 \boldsymbol{\mathcal{C}}^i\|_0 \leq \sum_{i=1}^4 \|\boldsymbol{\mathcal{C}}^i\|_0 \leq 16$. Necessarily, $\{\text{supp}(\boldsymbol{\mathcal{C}}^i)\}_{i=1}^r$ are pairwise disjoint.

Identifying the support constraint

Proposition

Let $\boldsymbol{Z}$ be the DFT, DCT-II or DST-II matrix. Then, for any EMD $\boldsymbol{Z} = \sum_{i=1}^{r} \boldsymbol{\mathcal{C}}^i$ with $\mathcal{C} \in \Gamma^{2,N/2}$, there exists σ such that: $\mathrm{supp}(\boldsymbol{\mathcal{C}}^i) \subseteq \boldsymbol{\mathcal{S}}^{\sigma(i)}$, *where $\{\boldsymbol{\mathcal{S}}^i\}_{i=1}^r$ are pairwise disjoint.*

$$\boldsymbol{DFT_4} = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{pmatrix}$$

1. If $\mathcal{C} \in \Gamma^{2,N/2}$ is an EMD of $\boldsymbol{DFT_4}$, then $16 = \|\boldsymbol{DFT_4}\|_0 = \|\sum_{i=1}^4 \boldsymbol{\mathcal{C}}^i\|_0 \leq \sum_{i=1}^4 \|\boldsymbol{\mathcal{C}}^i\|_0 \leq 16$. Necessarily, $\{\mathrm{supp}(\boldsymbol{\mathcal{C}}^i)\}_{i=1}^r$ are pairwise disjoint.
2. Only one possible partition $\{\boldsymbol{\mathcal{S}}^i\}_{i=1,2,3,4}$ of $\mathrm{supp}(\boldsymbol{DFT_4})$ such that $(\boldsymbol{DFT_4})_{|\boldsymbol{\mathcal{S}}^i}$ is of rank one.

Multilayer extension, with a butterfly sparsity structure

Given a matrix $\boldsymbol{Z}$ and a feasible set Σ of L-tuple of factors, define:

$$\text{find, if possible, } (\boldsymbol{X}_\ell) \in \Sigma \text{ such that } \boldsymbol{Z} = \boldsymbol{X}_{\boldsymbol{L}} \boldsymbol{X}_{\boldsymbol{L-1}} \ldots \boldsymbol{X}_{\boldsymbol{1}}. \qquad \text{(MEMF)}$$

Multilayer extension, with a butterfly sparsity structure

Given a matrix $\boldsymbol{Z}$ and a feasible set Σ of L-tuple of factors, define:

$$\text{find, if possible, } (\boldsymbol{X}_\ell) \in \Sigma \text{ such that } \boldsymbol{Z} = \boldsymbol{X_L}\boldsymbol{X_{L-1}}...\boldsymbol{X_1}. \qquad \text{(MEMF)}$$

Here: $\Sigma^{\text{fly}} := \{\boldsymbol{X_L}, ..., \boldsymbol{X_1} \text{ have supp included in the butterfly supports}\}$.

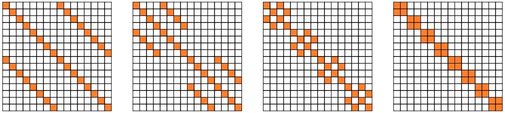


Figure: Butterfly supports: block diagonal + 2-sparse by row and by column.

Multilayer extension, with a butterfly sparsity structure

Given a matrix $\boldsymbol{Z}$ and a feasible set Σ of L-tuple of factors, define:

$$\text{find, if possible, } (\boldsymbol{X}_\ell) \in \Sigma \text{ such that } \boldsymbol{Z} = \boldsymbol{X}_{\boldsymbol{L}} \boldsymbol{X}_{\boldsymbol{L-1}} \ldots \boldsymbol{X}_{\boldsymbol{1}}. \qquad \text{(MEMF)}$$

Here: $\Sigma^{\text{fly}} := \{\boldsymbol{X}_{\boldsymbol{L}}, \ldots, \boldsymbol{X}_{\boldsymbol{1}} \text{ have supp included in the butterfly supports}\}$.

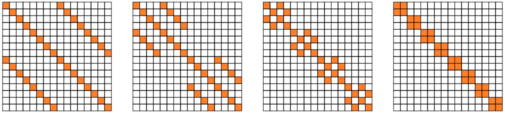


Figure: Butterfly supports: block diagonal + 2-sparse by row and by column.

Theorem

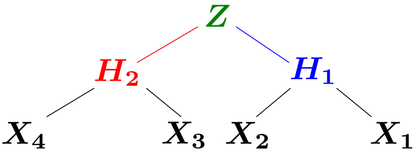
Let $\boldsymbol{Z} := \boldsymbol{X}_{\boldsymbol{L}} \boldsymbol{X}_{\boldsymbol{L-1}} \ldots \boldsymbol{X}_{\boldsymbol{1}}$ of size $N = 2^L$ where $\text{supp}(\boldsymbol{X}_{\boldsymbol{L}}), \ldots, \text{supp}(\boldsymbol{X}_{\boldsymbol{1}})$ are exactly the butterfly supports. Then, the factors $\boldsymbol{X}_{\boldsymbol{L}}, \ldots, \boldsymbol{X}_{\boldsymbol{1}}$ are the unique MEMF of $\boldsymbol{Z}$ in Σ^{fly}, up to scaling ambiguities.

Application: $\boldsymbol{Z}$ = DFT matrix of size $N = 2^L$.

A hierarchical factorization method

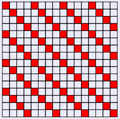
Consider $(\boldsymbol{X}_4, \boldsymbol{X}_3, \boldsymbol{X}_2, \boldsymbol{X}_1) \in \Sigma^{\text{fly}}$, and

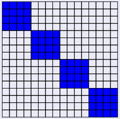
$$\boldsymbol{Z} = \boldsymbol{X}_4 \boldsymbol{X}_3 \boldsymbol{X}_2 \boldsymbol{X}_1.$$



Lemma

For any $(\boldsymbol{X}'_4, \boldsymbol{X}'_3, \boldsymbol{X}'_2, \boldsymbol{X}'_1) \in \Sigma^{\text{fly}}$*, we have:*

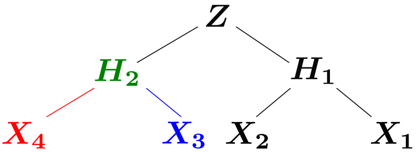
$\text{supp}(\boldsymbol{X}'_4 \boldsymbol{X}'_3) \subseteq$ 

$\text{supp}(\boldsymbol{X}'_2 \boldsymbol{X}'_1) \subseteq$ 

A hierarchical factorization method

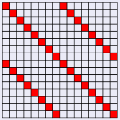
Consider $(\boldsymbol{X}_4, \boldsymbol{X}_3, \boldsymbol{X}_2, \boldsymbol{X}_1) \in \Sigma^{\text{fly}}$, and

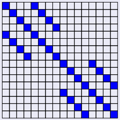
$$\boldsymbol{Z} = \boldsymbol{X}_4 \boldsymbol{X}_3 \boldsymbol{X}_2 \boldsymbol{X}_1.$$



Lemma

For any $(\boldsymbol{X}_4', \boldsymbol{X}_3', \boldsymbol{X}_2', \boldsymbol{X}_1') \in \Sigma^{\text{fly}}$, *we have:*

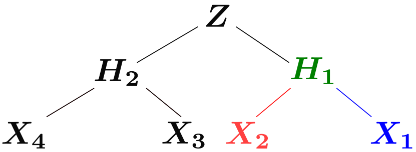
$\text{supp}(\boldsymbol{X}_4') \subseteq$ 

$\text{supp}(\boldsymbol{X}_3') \subseteq$ 

A hierarchical factorization method

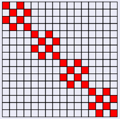
Consider $(\mathbf{X}_4, \mathbf{X}_3, \mathbf{X}_2, \mathbf{X}_1) \in \Sigma^{\text{fly}}$, and

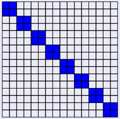
$$\mathbf{Z} = \mathbf{X}_4 \mathbf{X}_3 \mathbf{X}_2 \mathbf{X}_1.$$



Lemma

For any $(\mathbf{X}'_4, \mathbf{X}'_3, \mathbf{X}'_2, \mathbf{X}'_1) \in \Sigma^{\text{fly}}$, *we have:*

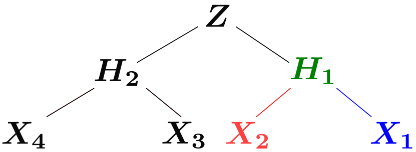
$\text{supp}(\mathbf{X}'_2) \subseteq$ 

$\text{supp}(\mathbf{X}'_1) \subseteq$ 

A hierarchical factorization method

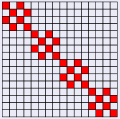
Consider $(\boldsymbol{X}_4, \boldsymbol{X}_3, \boldsymbol{X}_2, \boldsymbol{X}_1) \in \Sigma^{\text{fly}}$, and

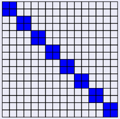
$$\boldsymbol{Z} = \boldsymbol{X}_4\boldsymbol{X}_3\boldsymbol{X}_2\boldsymbol{X}_1.$$



Lemma

For any $(\boldsymbol{X}'_4, \boldsymbol{X}'_3, \boldsymbol{X}'_2, \boldsymbol{X}'_1) \in \Sigma^{\text{fly}}$, *we have:*

$\text{supp}(\boldsymbol{X}'_2) \subseteq$ 

$\text{supp}(\boldsymbol{X}'_1) \subseteq$ 

This property of the butterfly supports is true for any number of layers, and any hierarchical tree structure.

Exact recovery of the multiple butterfly factors

Let $\boldsymbol{Z} := \boldsymbol{X_L X_{L-1} ... X_1}$ of size $N = 2^L$ where supp($\boldsymbol{X_L}$), ..., supp($\boldsymbol{X_1}$) are exactly the butterfly supports.

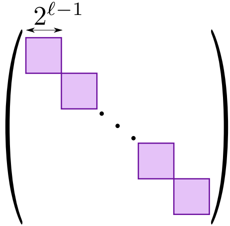
Algorithm Exact sparse recovery of $\boldsymbol{X_L}, ..., \boldsymbol{X_1}$ from $\boldsymbol{Z}$, up to rescaling.

Require: matrix $\boldsymbol{Z}$

$\boldsymbol{H} \leftarrow \boldsymbol{Z}$
for $\ell = L, ..., 1$ **do**
  $\mathcal{S} \leftarrow \varphi(\boldsymbol{B}^\ell, \boldsymbol{W}^{\ell-1})$
  for $i = 1, ..., r$ **do**
    $\boldsymbol{\mathcal{C}}^i \leftarrow \boldsymbol{H}_{|\boldsymbol{\mathcal{S}}^i}$
  end for
  $(\boldsymbol{X}'_\ell, \boldsymbol{H}^T) \leftarrow \varphi^{-1}(\mathcal{C})$
end for
return $\boldsymbol{X}'_L, ..., \boldsymbol{X}'_1$

$\boldsymbol{B}^\ell = \ell$-th butterfly support

$\boldsymbol{W}^{\ell-1} :=$



where  is a block full of 1s.

Recovery under noise: set $\boldsymbol{\mathcal{C}}^i$ as the best rank-one approximation of $\boldsymbol{H}_{|\boldsymbol{\mathcal{S}}^i}$.

Conclusion and discussion

Take-home message

1. Identifiability for **well-posedness** of sparse matrix factorization

Conclusion and discussion

Take-home message

1. Identifiability for **well-posedness** of sparse matrix factorization
2. Analysis of identifiability in multilinear inverse problems relies on the **lifting** approach

Conclusion and discussion

Take-home message

1. Identifiability for **well-posedness** of sparse matrix factorization
2. Analysis of identifiability in multilinear inverse problems relies on the **lifting** approach
3. Extension to the multilayer case via a **hierarchical** factorization method

Conclusion and discussion

Take-home message

1. Identifiability for **well-posedness** of sparse matrix factorization
2. Analysis of identifiability in multilinear inverse problems relies on the **lifting** approach
3. Extension to the multilayer case via a **hierarchical** factorization method

Future work

- Tighter conditions for fixed-support identifiability, to better understand identifiability of the support constraint.
- Identifiability in the multilayer case constrained by a *family* of sparsity patterns.

References

Choudhary, Sunav and Urbashi Mitra (2014). "Identifiability scaling laws in bilinear inverse problems". In: *arXiv preprint arXiv:1402.2637*.

Gillis, Nicolas (2020). *Nonnegative Matrix Factorization*. SIAM.

Le Magoarou, Luc (2016). "Matrices efficientes pour le traitement du signal et l'apprentissage automatique". PhD thesis. INSA de Rennes.

Le Magoarou, Luc and Rémi Gribonval (2016). "Flexible multilayer sparse approximations of matrices and applications". In: *IEEE Journal of Selected Topics in Signal Processing* 10.4, pp. 688–700.