



HAL
open science

Hierarchical Identifiability in Multi-layer Sparse Matrix Factorization

Léon Zheng, Elisa Riccietti, Rémi Gribonval

► **To cite this version:**

Léon Zheng, Elisa Riccietti, Rémi Gribonval. Hierarchical Identifiability in Multi-layer Sparse Matrix Factorization. 2021. hal-03362626v2

HAL Id: hal-03362626

<https://inria.hal.science/hal-03362626v2>

Preprint submitted on 11 Nov 2021 (v2), last revised 7 Oct 2022 (v6)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Hierarchical Identifiability in Multi-layer Sparse Matrix Factorization*

Léon Zheng^{‡†}, Elisa Riccietti[†], and Rémi Gribonval[†]

Abstract. Many well-known matrices \mathbf{Z} are associated to fast transforms corresponding to factorizations of the form $\mathbf{Z} = \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$, where each factor $\mathbf{X}^{(\ell)}$ is sparse and possibly structured. This paper investigates essential uniqueness of such factorizations. Our first main contribution is to prove that any $N \times N$ matrix having the so-called butterfly structure admits a unique factorization into J butterfly factors (where $N = 2^J$), and that the factors can be recovered by a hierarchical factorization method. This contrasts with existing approaches which fit the product of the butterfly factors to a given matrix via gradient descent. The proposed method can be applied in particular to retrieve the factorizations of the Hadamard or the Discrete Fourier Transform matrices of size 2^J . Computing such factorizations costs $\mathcal{O}(N^2)$, which is of the order of dense matrix-vector multiplication, while the obtained factorizations enable fast $\mathcal{O}(N \log N)$ matrix-vector multiplications. This hierarchical identifiability property relies on a simple identifiability condition in the two-layer and fixed-support setting that was recently established. While the butterfly structure corresponds to a fixed prescribed support for each factor, our second contribution is to obtain identifiability results with more general families of allowed sparsity patterns, taking into account permutation ambiguities when they are unavoidable. Typically, we show through the hierarchical paradigm that the butterfly factorization of the Discrete Fourier Transform matrix of size 2^J admits a unique sparse factorization into J factors, when enforcing only 2-sparsity by column and a block-diagonal structure on each factor.

Key words. Identifiability, matrix factorization, sparsity, hierarchical factorization, butterfly factorization

AMS subject classifications. 15A23, 65F50, 94A12

1. Introduction. Sparse matrix factorization with $J \geq 2$ factors is the problem of approximating a given matrix \mathbf{Z} by a product of J sparse factors $\mathbf{X}^{(J)} \mathbf{X}^{(J-1)} \dots \mathbf{X}^{(1)}$. Such a factorization is desired to reduce time and memory complexity for numerical methods involving the linear operator associated to \mathbf{Z} , e.g., in large-scale linear inverse problems [17, 18, 11, 12].

Sparsity constraints are usually encoded by a family of *allowed* supports, which force the factors to have some prescribed sparsity patterns. For instance, for $J = 2$, in the sparse coding problem [5, 4] with known dictionary $\mathbf{X}^{(2)}$, the factor $\mathbf{X}^{(1)}$ is constrained to be k -sparse by column, i.e., to have at most k nonzero entries per column.

This paper focuses on *identifiability* in the multi-layer sparse matrix factorization problem, i.e., uniqueness of the solution to this problem up to some unavoidable natural ambiguities. Identifiability results given in this paper are based on the general framework introduced in the companion paper [20], and they typically take the following form:

Informal Theorem 1.1. *Let \mathbf{Z} be a matrix, and Ω be a family of sparsity patterns. If a certain condition on Ω is satisfied, then \mathbf{Z} admits a unique sparse factorization $\mathbf{Z} = \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$, up to some unavoidable ambiguities, when the constraints on the factors are described by Ω .*

*Preprint.

Funding: This project was supported in part by the AllegroAssai ANR project ANR-19-CHIA-0009.

[†]Université de Lyon, ENS de Lyon, UCBL, CNRS, Inria, LIP, F-69342, LYON Cedex 07, France. (leon.zheng@ens-lyon.fr, elisa.riccietti@ens-lyon.fr, remi.gribonval@inria.fr). [‡]valeo.ai, Paris, France.

Understanding under which conditions the sparse matrix factorization problem is well-posed, in the sense that it admits a unique solution, is still very much an open question, but is the key to derive methods that are *guaranteed* to compute a good sparse approximation of the target matrix.

Typically, heuristic methods based on iterative first-order optimization [3, 12] do not have this kind of guarantees, and yield poor approximations of the target matrix, unless the parameters are well initialized. This kind of methods is applied to the following formulation of the sparse matrix factorization problem:

$$(1.1) \quad \min_{\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}} \|\mathbf{Z} - \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}\|_F, \quad \text{where } \|\cdot\|_F \text{ denotes the Frobenius norm,}$$

with the constraint that the factors $\mathbf{X}^{(\ell)}$, $1 \leq \ell \leq J$, are sparse, possibly with structured constraints on their supports. For example, k -sparsity can be imposed globally on $\mathbf{X}^{(\ell)}$, or on each of its rows, or each column, or both. Proximal gradient algorithms [12] can for instance be employed to explore the family of sparsity patterns, and find an allowed sparsity pattern that yields the smallest approximation error. To overcome the difficulties of using this technique in the multi-layer setting, a hierarchical factorization method has been proposed [12], in which the target matrix is iteratively factorized into two factors, until the desired number of sparse factors J is obtained. However, recent works have shown some important pitfalls of this hierarchical paradigm [8]: it is possible that an intermediate factor obtained at a certain level of the hierarchy cannot be further factorized into factors with admissible sparsity patterns, which yields a poor overall performance of the hierarchical method.

This paper shows that when \mathbf{Z} admits an *exact* sparse factorization $\mathbf{Z} = \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$, such a hierarchical method can be guaranteed to recover the sparse factors $(\mathbf{X}^{(\ell)})_{\ell=1}^J$ up to some unavoidable ambiguities, provided that the intermediate sparsity constraint at each level of the hierarchy is well-chosen. In this case, the pitfall mentioned above can be avoided.

Specifically, our first main contribution is to show that the so-called *butterfly* structure [11, 3], which appears naturally in many fast transforms such as the Discrete Fourier Transform (DFT) or the Hadamard Transform, is a good choice of *fixed* structured support constraint that ensures such a *hierarchical identifiability* (see [Theorem 3.10](#)). We derive from our analysis a natural associated algorithm for the recovery of the sparse butterfly factors $\mathbf{X}^{(\ell)}$, $1 \leq \ell \leq J$, from their product $\mathbf{Z} = \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$ (see [Algorithm 3.1](#)). In contrast to existing approaches for butterfly factorization, which are based on iterative gradient descent [3, 12], the proposed algorithm has bounded complexity and is endowed with exact recovery guarantees.

The proof of the hierarchical identifiability property relies on a simple sufficient condition presented in the companion paper [20] for identifiability in exact sparse matrix factorization with $J = 2$ factors and *fixed* supports, i.e., when the supports of the two factors are known. This condition is based on the *lifting* principle commonly considered in multilinear inverse problems [1, 2, 13, 16, 14, 15], which, in the context of matrix factorization into two factors, proposes to represent a pair of factors $(\mathbf{X}^{(2)}, \mathbf{X}^{(1)})$ by the tuple of rank-one matrices $(\mathbf{C}^i)_i$, which are the outer products between columns of $\mathbf{X}^{(2)}$ and rows of $\mathbf{X}^{(1)}$ of the same index i .

To go beyond the fixed-support setting, our second contribution is to show some identifiability results in the case where the supports of the factors are no longer known, but belong to a larger family of sparsity patterns, e.g., k -sparsity by column and/or by row. Since these

families are invariant to column permutation, such permutation ambiguities must be taken into account in the uniqueness property [20]. We illustrate this kind of extended identifiability properties on the factorization into two factors of some ubiquitous matrices like the DFT, Discrete Cosine Transform (DCT) or Discrete Sine Transform (DST) matrices, for some well-chosen family of sparsity patterns. Remarkably, uniqueness fails to hold for the Hadamard matrix, under the same sparsity constraint. In the multi-layer setting, we show that the DFT matrix of size 2^J admits the butterfly factorization as the unique sparse factorization into J factors, also when enforcing *only* 2-sparsity by column and a block-diagonal structure on each factor, instead of the specific butterfly structure (see [Theorem 4.15](#)). Again, this factorization can be retrieved through a hierarchical factorization method, when choosing carefully the family of sparsity patterns which encodes the constraint at each hierarchical level.

Summary. The main contributions of this paper are the following ones:

1. [Theorem 3.10](#) shows that enforcing the butterfly structure on the J sparse factors is sufficient to ensure a hierarchical identifiability property, meaning that we can recover (up to scaling ambiguities) sparse factors $(\mathbf{X}^{(\ell)})_{\ell=1}^J$ from $\mathbf{Z} := \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$, using [Algorithm 3.1](#) which has a time complexity of only $\mathcal{O}(N^2)$ where N is the size of \mathbf{Z} . This is remarkably of the same order of magnitude as the complexity of matrix-vector multiplication, and can be performed only once to enable $\mathcal{O}(N \log N)$ matrix-vector multiplications with the resulting factored representation of \mathbf{Z} .
2. [Theorem 4.5](#) establishes uniqueness of the sparse factorization into two factors of the DFT, DCT and DST matrices, when enforcing $\frac{N}{2}$ -sparsity by column on the left factor, and 2-sparsity by row on the right factor, which contrasts with the non-identifiability of the sparse factorization of the Hadamard matrix when enforcing the same constraint.
3. [Theorem 4.15](#) shows that a relaxation of the sparsity constraint to a block-diagonal structure and to 2-sparsity by column on each factor yields a stronger identifiability result for the butterfly factorization of the DFT matrix of size 2^J into J factors.

The paper is organized as follows: [section 2](#) recalls the general framework used to analyze identifiability in the two-layer setting introduced in the companion paper; [section 3](#) describes how the butterfly structure can ensure the hierarchical identifiability of the sparse factors from their product; [section 4](#) illustrates how we can prove identifiability results when the sparsity constraint is described by a family of prescribed sparsity patterns; [section 5](#) discusses perspectives of this work for algorithmic methods. An annex gathers technical proofs.

Notations. The set of integers $\{1, \dots, n\}$ is denoted $\llbracket n \rrbracket$. The *support* of a matrix $\mathbf{M} \in \mathbb{C}^{m \times n}$ of size $m \times n$ is the set of indices $\text{supp}(\mathbf{M}) \subseteq \llbracket m \rrbracket \times \llbracket n \rrbracket$ of nonzero entries. It is identified by abuse to the set of binary matrices $\mathbb{B}^{m \times n} := \{0, 1\}^{m \times n}$. Depending on the context, a matrix support can be seen as a set of indices, or a binary matrix with only nonzero entries for indices in this set. The cardinality of $\text{supp}(\mathbf{M})$ is also known as the ℓ_0 -norm of \mathbf{M} , denoted $\|\cdot\|_0$. The *column support*, denoted $\text{colsupp}(\mathbf{M})$, is the subset of indices $i \in \llbracket r \rrbracket$ such that the i -th column of \mathbf{M} , denoted \mathbf{M}_i , is nonzero. Similarly, the *row support* of a matrix \mathbf{M} is the column support of its transpose \mathbf{M}^\top , and is denoted $\text{rowsupp}(\mathbf{M})$. The notation $\mathbf{M}^{(i)}$ denotes the i -th matrix in a collection. The entry of \mathbf{M} indexed by (k, l) is $\mathbf{M}_{k,l}$. The identity matrix of size n is denoted \mathbf{I}_n . The Kronecker product [19] between \mathbf{A} and \mathbf{B} is written $\mathbf{A} \otimes \mathbf{B}$. We recall:

$$(1.2) \quad (\mathbf{A} \otimes \mathbf{C})(\mathbf{B} \otimes \mathbf{D}) = \mathbf{AB} \otimes \mathbf{CD}.$$

2. A general framework for identifiability in two-layer sparse matrix factorization. In order to show hierarchical identifiability results, we rely on tools developed in the companion paper [20] for the analysis of identifiability in two-layer sparse matrix factorization. *This section is almost a word-to-word restatement of the main results of [20], and can be skipped by a reader who already read this companion paper.*

Given a matrix $\mathbf{Z} \in \mathbb{C}^{m \times n}$, and a subset of pairs of factors $\Sigma \subseteq \mathbb{C}^{m \times r} \times \mathbb{C}^{n \times r}$, the so-called *exact matrix factorization* (EMF) problem with two factors of \mathbf{Z} in Σ is:

$$(2.1) \quad \text{find if possible } (\mathbf{X}, \mathbf{Y}) \in \Sigma \text{ such that } \mathbf{Z} = \mathbf{X}\mathbf{Y}^\top.$$

We are interested in the particular problem variation where the constraint set Σ encodes some chosen sparsity patterns for the factorization. For a given binary matrix $\mathbf{S} \in \mathbb{B}^{m \times r}$ associated to a sparsity pattern, denote

$$(2.2) \quad \Sigma_{\mathbf{S}} := \{\mathbf{M} \in \mathbb{C}^{m \times r} \mid \text{supp}(\mathbf{M}) \subseteq \text{supp}(\mathbf{S})\},$$

which is the set of matrices with a sparsity pattern included in \mathbf{S} . A pair of sparsity patterns is written $\mathbf{S} := (\mathbf{S}^L, \mathbf{S}^R)$, where \mathbf{S}^L and \mathbf{S}^R are the left and right sparsity patterns respectively. By abuse, \mathbf{S}^L and \mathbf{S}^R are respectively referred to as a left and a right (allowed) support. Given any family $\Omega \subset \mathbb{B}^{m \times r} \times \mathbb{B}^{n \times r}$ of such pairs of allowed supports, denote

$$(2.3) \quad \Sigma_{\Omega} := \bigcup_{\mathbf{S} \in \Omega} \Sigma_{\mathbf{S}}, \quad \text{with } \mathbf{S} := (\mathbf{S}^L, \mathbf{S}^R), \text{ and } \Sigma_{\mathbf{S}} := \Sigma_{\mathbf{S}^L} \times \Sigma_{\mathbf{S}^R}.$$

Since the support of a matrix is unchanged under arbitrary rescaling of its columns, Σ_{Ω} is invariant by column scaling for any family Ω . The framework covers some classical families Ω of structured sparse supports, like global sparsity, sparsity by column and/or row. These classical families are also invariant to permutations of columns, and will be referred to as families *stable* by permutation. Hence, uniqueness of a solution to (2.1) with such sparsity constraints will always be considered up to scaling and permutation ambiguities. Denote \mathcal{P}_r , \mathcal{D}_r respectively as the group of permutation and invertible diagonal matrices of size $r \times r$, and $\mathcal{G}_r := \{\mathbf{D}\mathbf{P} \mid \mathbf{D} \in \mathcal{D}_r, \mathbf{P} \in \mathcal{P}_r\}$ as the group of *generalized permutation* matrices.

Definition 2.1 (PS-uniqueness of an EMF in Σ [20, Definition 2.5]). *For any set Σ of pairs of factors, the pair $(\mathbf{X}, \mathbf{Y}) \in \Sigma$ is the PS-unique EMF of $\mathbf{Z} := \mathbf{X}\mathbf{Y}^\top$ in Σ , if any solution $(\mathbf{X}', \mathbf{Y}')$ to (2.1) with \mathbf{Z} and Σ is equivalent to (\mathbf{X}, \mathbf{Y}) , written $(\mathbf{X}', \mathbf{Y}') \sim (\mathbf{X}, \mathbf{Y})$, in the sense that there exists a generalized permutation matrix $\mathbf{G} \in \mathcal{G}_r$ such that $(\mathbf{X}', \mathbf{Y}') = (\mathbf{X}\mathbf{G}, \mathbf{Y}(\mathbf{G}^{-1})^\top)$.*

For any set Σ of pairs of factors, the set of all pairs $(\mathbf{X}, \mathbf{Y}) \in \Sigma$ such that (\mathbf{X}, \mathbf{Y}) is the PS-unique EMF of $\mathbf{Z} := \mathbf{X}\mathbf{Y}^\top$ in Σ is denoted $\mathcal{U}(\Sigma)$. In other words, we define:

$$(2.4) \quad \mathcal{U}(\Sigma) := \{(\mathbf{X}, \mathbf{Y}) \in \Sigma \mid \forall (\mathbf{X}', \mathbf{Y}') \in \Sigma, \mathbf{X}'\mathbf{Y}'^\top = \mathbf{X}\mathbf{Y}^\top \implies (\mathbf{X}', \mathbf{Y}') \sim (\mathbf{X}, \mathbf{Y})\}.$$

In order to characterize the set $\mathcal{U}(\Sigma_{\Omega})$ for Σ_{Ω} defined as in (2.3) with Ω any family that is *stable* by permutation, two non-degeneration properties were established for a pair of factors [20], i.e., two necessary conditions for identifiability, involving their so-called *column support*.

Define the set of pairs of factors with *identical* and *maximal* column supports in Σ_Ω as

$$(2.5) \quad \text{IC}_\Omega := \{(\mathbf{X}, \mathbf{Y}) \in \Sigma_\Omega \mid \text{colsupp}(\mathbf{X}) = \text{colsupp}(\mathbf{Y})\}.$$

$$(2.6) \quad \text{MC}_\Omega := \{(\mathbf{X}, \mathbf{Y}) \in \Sigma_\Omega \mid \forall \mathbf{S} \in \Omega \text{ such that } (\mathbf{X}, \mathbf{Y}) \in \Sigma_{\mathbf{S}}, \text{colsupp}(\mathbf{X}) = \text{colsupp}(\mathbf{S}^L) \\ \text{and } \text{colsupp}(\mathbf{Y}) = \text{colsupp}(\mathbf{S}^R)\}.$$

Lemma 2.2 ([20, Lemma 2.11]). *For any family of pairs of supports Ω : $\mathcal{U}(\Sigma_\Omega) \subseteq \text{IC}_\Omega \cap \text{MC}_\Omega$.*

As the matrix product $\mathbf{X}\mathbf{Y}^\top$ can be decomposed into the sum of rank-one matrices $\sum_{i=1}^r \mathbf{X}_i \mathbf{Y}_i^\top$, the so-called lifting procedure [1, 15] suggests to represent a pair (\mathbf{X}, \mathbf{Y}) by its r -tuple of so-called *rank-one contributions*. With this lifting approach, each support constraint $\mathbf{S} = (\mathbf{S}^L, \mathbf{S}^R)$ is represented by the r -tuple of rank-one support constraints $\mathcal{S} = \varphi(\mathbf{S}^L, \mathbf{S}^R)$. Thus, when (\mathbf{X}, \mathbf{Y}) follows a sparsity structure given by Ω , i.e., belongs to $\Sigma_\Omega \subseteq \mathbb{C}^{m \times r} \times \mathbb{C}^{n \times r}$, the r -tuple of rank-one matrices $\varphi(\mathbf{X}, \mathbf{Y}) \in (\mathbb{C}^{m \times n})^r$ belongs to the set:

$$(2.7) \quad \Gamma_\Omega := \bigcup_{\mathcal{S} \in \varphi(\Omega)} \Gamma_{\mathcal{S}}, \quad \text{with } \mathcal{S} := (\mathbf{S}^1, \dots, \mathbf{S}^r), \quad \text{and}$$

$$(2.8) \quad \Gamma_{\mathcal{S}} := \{(\mathbf{C}^i)_{i=1}^r \mid \forall i \in \llbracket r \rrbracket, \text{rank}(\mathbf{C}^i) \leq 1, \text{supp}(\mathbf{C}^i) \subseteq \text{supp}(\mathbf{S}^i)\} \subseteq (\mathbb{C}^{m \times n})^r.$$

Denote the operator which sums the r matrices of a tuple \mathcal{C} as:

$$(2.9) \quad \mathcal{A} : \mathcal{C} = (\mathbf{C}^i)_{i=1}^r \mapsto \sum_{i=1}^r \mathbf{C}^i$$

Definition 2.3 (P-uniqueness of an EMD in Γ [20, Definition 4.5]). *For any set $\Gamma \subseteq (\mathbb{C}^{m \times n})^r$ of r -tuples of rank-one matrices, the r -tuple $\mathcal{C} \in \Gamma$ is the P-unique exact matrix decomposition (EMD) of $\mathbf{Z} := \mathcal{A}(\mathcal{C})$ in Γ if, for any $\mathcal{C}' \in \Gamma$ such that $\mathcal{A}(\mathcal{C}') = \mathbf{Z}$, we have $\mathcal{C}' \sim \mathcal{C}$, in the sense that the tuples $\mathcal{C} := (\mathbf{C}^i)_{i=1}^r$, $\mathcal{C}' := (\mathbf{C}'^i)_{i=1}^r$ are equal up to a permutation of the index i .*

The set of all r -tuples $\mathcal{C} \in \Gamma$ such that \mathcal{C} is the P-unique EMD of $\mathbf{Z} := \mathcal{A}(\mathcal{C})$ in Γ is denoted $\mathcal{U}(\Gamma)$, where the notation $\mathcal{U}(\cdot)$ has been slightly abused:

$$(2.10) \quad \mathcal{U}(\Gamma) := \{\mathcal{C} \in \Gamma \mid \forall \mathcal{C}' \in \Gamma, \mathcal{A}(\mathcal{C}) = \mathcal{A}(\mathcal{C}') \implies \mathcal{C} \sim \mathcal{C}'\}.$$

Theorem 2.4 ([20, Theorem 4.8, Corollary 4.10]). *For any family of pairs of supports Ω stable by permutation, and any pair of factors (\mathbf{X}, \mathbf{Y}) :*

$$(\mathbf{X}, \mathbf{Y}) \in \mathcal{U}(\Sigma_\Omega) \iff \varphi(\mathbf{X}, \mathbf{Y}) \in \mathcal{U}(\Gamma_\Omega) \text{ and } (\mathbf{X}, \mathbf{Y}) \in \text{IC}_\Omega \cap \text{MC}_\Omega.$$

In the fixed-support setting, for any pair of supports \mathbf{S} , and any pair of factors (\mathbf{X}, \mathbf{Y}) , denoting $\mathcal{S} := \varphi(\mathbf{S})$, we have:

$$(\mathbf{X}, \mathbf{Y}) \in \mathcal{U}(\Sigma_{\mathbf{S}}) \iff \varphi(\mathbf{X}, \mathbf{Y}) \in \mathcal{U}(\Gamma_{\mathcal{S}}) \text{ and } (\mathbf{X}, \mathbf{Y}) \in \text{IC}_{\mathbf{S}} \cap \text{MC}_{\mathbf{S}},$$

where $\text{IC}_{\mathbf{S}}$ and $\text{MC}_{\mathbf{S}}$ are the specialization of IC_Ω and MC_Ω defined by (2.5) and (2.6) to the case where Ω is reduced to a singleton \mathbf{S} :

$$(2.11) \quad \text{IC}_{\mathbf{S}} := \{(\mathbf{X}, \mathbf{Y}) \in \Sigma_{\mathbf{S}} \mid \text{colsupp}(\mathbf{X}) = \text{colsupp}(\mathbf{Y})\},$$

$$(2.12) \quad \text{MC}_{\mathbf{S}} := \{(\mathbf{X}, \mathbf{Y}) \in \Sigma_{\mathbf{S}} \mid \text{colsupp}(\mathbf{X}) = \text{colsupp}(\mathbf{S}^L), \text{colsupp}(\mathbf{Y}) = \text{colsupp}(\mathbf{S}^R)\}.$$

The companion paper [20] provides a simple sufficient condition on the r -tuple of rank-one supports \mathcal{S} for P-uniqueness of an EMD in $\Gamma_{\mathcal{S}}$.

Lemma 2.5 ([20, Lemma 5.1]). *Let \mathcal{S} be an r -tuple of rank-one supports. Then, uniform P-uniqueness of EMD in $\Gamma_{\mathcal{S}}$ holds, i.e., $\mathcal{U}(\Gamma_{\mathcal{S}}) = \Gamma_{\mathcal{S}}$, if, and only if, the rank-one supports $\{\mathcal{S}^i\}_{i=1}^r$ are pairwise disjoint.*

Consequently, when a pair of supports $\mathbf{S} := (\mathbf{S}^L, \mathbf{S}^R)$ is such that $\varphi(\mathbf{S})$ has disjoint rank-one supports, almost every pair of factors (\mathbf{X}, \mathbf{Y}) is identifiable for the EMF of $\mathbf{Z} = \mathbf{X}\mathbf{Y}^\top$ in $\Sigma_{\mathbf{S}}$. In fact, a stronger identifiability than PS-uniqueness is verified in this case:

Definition 2.6 (S-uniqueness of an EMF in Σ [20, Definition 5.2]). *For any set Σ of pairs of factors, the pair $(\mathbf{X}, \mathbf{Y}) \in \Sigma$ is the S-unique EMF of $\mathbf{Z} := \mathbf{X}\mathbf{Y}^\top$ in Σ , if any solution $(\mathbf{X}', \mathbf{Y}')$ to (2.1) with \mathbf{Z} and Σ is equivalent to (\mathbf{X}, \mathbf{Y}) up to scaling ambiguities only, in the sense that there exists an invertible diagonal matrix $\mathbf{D} \in \mathcal{D}_r$ such that $(\mathbf{X}', \mathbf{Y}') = (\mathbf{X}\mathbf{D}, \mathbf{Y}\mathbf{D}^{-1})$.*

The set of all pairs $(\mathbf{X}, \mathbf{Y}) \in \Sigma$ such that (\mathbf{X}, \mathbf{Y}) is the S-unique EMF of $\mathbf{Z} := \mathbf{X}\mathbf{Y}^\top$ in Σ is denoted $\mathcal{U}_{\mathbf{S}}(\Sigma)$.

Proposition 2.7 ([20, Proposition 5.3]). *Suppose that $\mathbf{S} := (\mathbf{S}^L, \mathbf{S}^R)$ is such that the tuple $\varphi(\mathbf{S})$ has disjoint rank-one supports. Then: $\mathcal{U}_{\mathbf{S}}(\Sigma_{\mathbf{S}}) = \text{IC}_{\mathbf{S}} \cap \text{MC}_{\mathbf{S}}$.*

We will see in the following that this condition of disjoint rank-one supports will be verified when considering a so-called butterfly structure on the sparse factors.

3. Hierarchical identifiability with butterfly structure. The first main contribution of the paper is to show the hierarchical identifiability of the butterfly factors, i.e., we establish some identifiability results in the multi-layer sparse matrix factorization in a specific setting where the sparse factors are constrained to have the so-called *butterfly supports*.

3.1. Properties of the butterfly structure. Let us give a formal introduction of the butterfly structure and its important properties used to establish identifiability results.

Definition 3.1 (Butterfly supports). *The butterfly supports of size $N = 2^J$ are the J -tuple of supports $\mathbf{S}_{\text{bf}} := (\mathbf{S}_{\text{bf}}^J, \dots, \mathbf{S}_{\text{bf}}^1) \in (\mathbb{B}^{N \times N})^J$ defined by:*

$$(3.1) \quad \mathbf{S}_{\text{bf}}^\ell := \mathbf{I}_{N/2^\ell} \otimes \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \otimes \mathbf{I}_{2^{\ell-1}}, \quad 1 \leq \ell \leq J.$$

Figure 1 is an illustration of the butterfly supports. They are 2-regular [8], i.e., they have at most 2 nonzero entries per row and per column, and they are also block-diagonal, including the leftmost factor if we consider a single block that covers the matrix entirely.

Definition 3.2 (Butterfly structure). *We say that a matrix \mathbf{Z} of size $N = 2^J$ admits a butterfly structure if it can be factorized into J factors $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)})$ which have a support included in the butterfly supports $\mathbf{S}_{\text{bf}} := (\mathbf{S}_{\text{bf}}^J, \dots, \mathbf{S}_{\text{bf}}^1)$, in the sense that:*

$$(3.2) \quad \mathbf{Z} = \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}, \quad \text{with} \quad (\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{\mathbf{S}_{\text{bf}}} := \Sigma_{\mathbf{S}_{\text{bf}}^J} \times \dots \times \Sigma_{\mathbf{S}_{\text{bf}}^1},$$

where the notation $\Sigma_{\mathbf{S}}$ for any given subset of indices \mathbf{S} has been defined by (2.2).

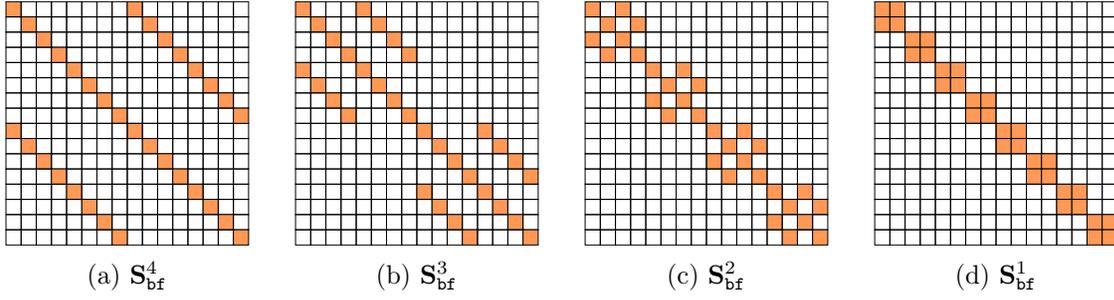


Figure 1: Butterfly supports of size $N = 16$. Nonzero entries are in color, and zero entries are in white.

Such a structure is involved in many fast linear transforms [3] such as the Hadamard, DST and DCT matrices of size 2^J , or as we detail here in the DFT matrix.

Example 3.3 (Butterfly factorization of the DFT matrix [3]). Consider the DFT matrix of size $N \times N$ with $N = 2^J$ denoted as \mathbf{DFT}_N , defined by:

$$(3.3) \quad \mathbf{DFT}_N := (\omega_N^{(k-1)(l-1)})_{k,l \in \llbracket N \rrbracket}, \quad \text{where } \omega_N := e^{-i\frac{2\pi}{N}}.$$

The butterfly factorization of \mathbf{DFT}_N [3] relies on the recursive relation

$$(3.4) \quad \mathbf{DFT}_N = \mathbf{B}_N \begin{pmatrix} \mathbf{DFT}_{N/2} & 0 \\ 0 & \mathbf{DFT}_{N/2} \end{pmatrix} \mathbf{P}_N,$$

where $\mathbf{P}_N \in \mathbb{B}^{N \times N}$ is the permutation matrix which sorts the odd indices, then the even indices, e.g., for $N = 4$, it permutes $(1, 2, 3, 4)$ to $(1, 3, 2, 4)$, and

$$(3.5) \quad \mathbf{B}_N := \begin{pmatrix} \mathbf{I}_{N/2} & \mathbf{A}_{N/2} \\ \mathbf{I}_{N/2} & -\mathbf{A}_{N/2} \end{pmatrix},$$

with $\mathbf{A}_{N/2}$ the diagonal matrix with diagonal entries $1, \omega_N, \omega_N^2, \dots, \omega_N^{\frac{N}{2}-1}$. Applying recursively (3.4) to the block $\mathbf{DFT}_{N/2}$, we obtain the butterfly factorization of \mathbf{DFT}_N :

$$(3.6) \quad \mathbf{DFT}_N = \mathbf{F}^{(J)} \dots \mathbf{F}^{(1)} \mathbf{R}_N, \quad \text{where } \mathbf{F}^{(\ell)} := \mathbf{I}_{N/2^\ell} \otimes \mathbf{B}_{2^\ell}, \quad 1 \leq \ell \leq J,$$

and $\mathbf{R}_N \in \mathbb{B}^{N \times N}$ is the so-called *bit-reversal* permutation matrix, defined by:

$$(3.7) \quad \mathbf{R}_N := \mathbf{Q}^{(1)} \mathbf{Q}^{(2)} \dots \mathbf{Q}^{(J)}, \quad \text{where } \mathbf{Q}^{(\ell)} := \mathbf{I}_{N/2^\ell} \otimes \mathbf{P}_{2^\ell}, \quad 1 \leq \ell \leq J.$$

Given $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{\mathbf{S}_{\text{bf}}}$ where \mathbf{S}_{bf} is the tuple of butterfly supports of size 2^J , we show that the partial product of any consecutive factors $\mathbf{X}^{(q)} \dots \mathbf{X}^{(p)}$ ($1 \leq p \leq q \leq J$) has a very precise structure as detailed in the following lemma. Figure 2 illustrates this structure on some examples of partial products for the case of the butterfly supports of size $N = 16$.

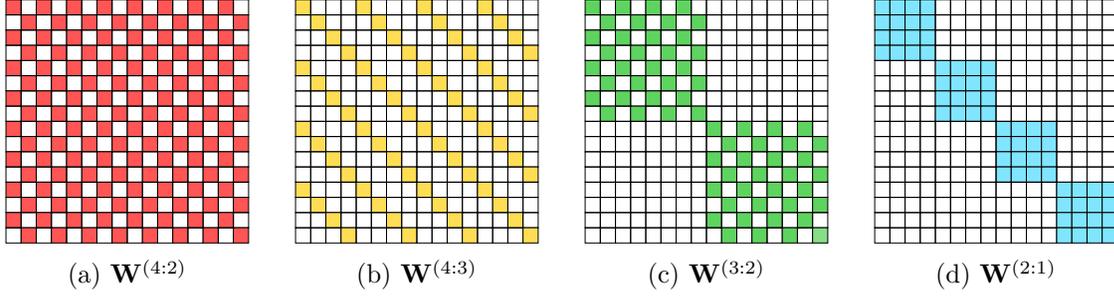


Figure 2: Examples of supports $\mathbf{W}^{(q;p)}$ defined by (3.8) ($1 \leq p \leq q \leq 4$) of size $N \times N$ with $N = 16$. Nonzero entries are in color, and zero entries are in white.

Lemma 3.4. Let $\mathbf{S}_{\text{bf}} := (\mathbf{S}_{\text{bf}}^J, \dots, \mathbf{S}_{\text{bf}}^1)$ be the butterfly supports of size $N = 2^J$. Then, for any tuple $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{\mathbf{S}_{\text{bf}}}$, for any $1 \leq p \leq q \leq J$: $\text{supp}(\mathbf{X}^{(q)} \dots \mathbf{X}^{(p)}) \subseteq \mathbf{W}^{(q;p)}$, where

$$(3.8) \quad \mathbf{W}^{(q;p)} := \mathbf{I}_{N/2^q} \otimes \mathbf{V}^{q;p} = \begin{pmatrix} \mathbf{V}^{q;p} & & 0 \\ & \ddots & \\ 0 & & \mathbf{V}^{q;p} \end{pmatrix} \in \mathbb{B}^{N \times N}, \text{ and}$$

$$(3.9) \quad \mathbf{V}^{q;p} := \mathbf{U}_{2^{q-p+1}} \otimes \mathbf{I}_{2^{p-1}} = \begin{pmatrix} \mathbf{I}_{2^{p-1}} & \dots & \mathbf{I}_{2^{p-1}} \\ \vdots & & \vdots \\ \mathbf{I}_{2^{p-1}} & \dots & \mathbf{I}_{2^{p-1}} \end{pmatrix} \in \mathbb{B}^{2^q \times 2^q},$$

denoting, for any n , $\mathbf{U}_n \in \mathbb{B}^{n \times n}$ as the binary matrix full of ones.

Remark 3.5. We have $\mathbf{W}^{(\ell;\ell)} = \mathbf{S}_{\text{bf}}^\ell$ for $1 \leq \ell \leq J$. Moreover, viewing matrix supports as binary matrices, one can verify that $\mathbf{W}^{(q;p)} = \mathbf{S}_{\text{bf}}^q \dots \mathbf{S}_{\text{bf}}^p$ for $1 \leq p \leq q \leq J$. Also, by (3.9), $\mathbf{V}^{q;p}$ is symmetric, i.e., $\mathbf{V}^{q;p} = (\mathbf{V}^{q;p})^\top$. By (3.8), $\mathbf{W}^{(q;p)}$ is block diagonal where each block is $\mathbf{V}^{q;p}$, so $\mathbf{W}^{(q;p)}$ is also symmetric. Finally, we remark that $\mathbf{W}^{(q;p)}$ and $\mathbf{V}^{q;p}$ are both 2^{q-p+1} -sparse by column.

The proof of this lemma follows from the above remark and the following lemma.

Lemma 3.6. Given two matrix supports $\mathbf{S}^2 \in \mathbb{B}^{m \times r}$ and $\mathbf{S}^1 \in \mathbb{B}^{r \times n}$, for any $(\mathbf{X}^{(2)}, \mathbf{X}^{(1)}) \in \Sigma_{\mathbf{S}^2} \times \Sigma_{\mathbf{S}^1}$, we have $\text{supp}(\mathbf{X}^{(2)}\mathbf{X}^{(1)}) \subseteq \text{supp}(\mathbf{S}^2\mathbf{S}^1)$.

Proof. If $(i, j) \notin \text{supp}(\mathbf{S}^2\mathbf{S}^1)$ then $0 = (\mathbf{S}^2\mathbf{S}^1)_{i,j} = \sum_{k=1}^r \mathbf{S}^2_{i,k} \mathbf{S}^1_{k,j}$. As $\mathbf{S}^2, \mathbf{S}^1$ are binary matrices, this means that for each $k \in \llbracket r \rrbracket$, $\mathbf{S}^2_{i,k} = 0$ or $\mathbf{S}^1_{k,j} = 0$. Since $\text{supp}(\mathbf{X}^{(2)}) \subseteq \mathbf{S}^2$ and $\text{supp}(\mathbf{X}^{(1)}) \subseteq \mathbf{S}^1$, we have $\mathbf{X}^{(2)}_{i,k} = 0$ or $\mathbf{X}^{(1)}_{k,j} = 0$, for each $k \in \llbracket r \rrbracket$. This yields: $(\mathbf{X}^{(2)}\mathbf{X}^{(1)})_{i,j} = \sum_{k=1}^r \mathbf{X}^{(2)}_{i,k} \mathbf{X}^{(1)}_{k,j} = 0$. Hence, $(i, j) \notin \text{supp}(\mathbf{X}^{(2)}\mathbf{X}^{(1)})$. By contraposition this establishes the result. \blacksquare

Proof of Lemma 3.4. We start by the case where $q = J$. Let us show by backward induction that $\text{supp}(\mathbf{X}^{(J)} \dots \mathbf{X}^{(\ell)}) \subseteq \mathbf{V}^{J,\ell}$ for any $1 \leq \ell \leq J$. This is true for $\ell = J$, because by (3.1) and (3.9), $\text{supp}(\mathbf{X}^{(J)}) \subseteq \mathbf{S}_{\text{bf}}^J = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \otimes \mathbf{I}_{2^{J-1}} = \mathbf{V}^{J,J}$. Let $2 \leq \ell \leq J$, and suppose that $\text{supp}(\mathbf{X}^{(J)} \dots \mathbf{X}^{(\ell)}) \subseteq \mathbf{V}^{J,\ell}$, which is the matrix full of blocks $\mathbf{I}_{2^{\ell-1}}$. Since $\text{supp}(\mathbf{X}^{(\ell-1)}) \subseteq \mathbf{S}_{\text{bf}}^{\ell-1}$,

by Lemma 3.6, we have $\text{supp}(\mathbf{X}^{(J)} \dots \mathbf{X}^{(\ell)} \mathbf{X}^{(\ell-1)}) \subseteq \text{supp}(\mathbf{V}^{J,\ell} \mathbf{S}_{\text{bf}}^{\ell-1})$. But the support $\mathbf{S}_{\text{bf}}^{\ell-1}$ is block diagonal with blocks of size $2^{\ell-1} \times 2^{\ell-1}$ equal to $\mathbf{U}_2 \otimes \mathbf{I}_{2^{\ell-2}} = \begin{bmatrix} \mathbf{I}_{2^{\ell-2}} & \mathbf{I}_{2^{\ell-2}} \\ \mathbf{I}_{2^{\ell-2}} & \mathbf{I}_{2^{\ell-2}} \end{bmatrix}$, so:

$$\begin{aligned} \mathbf{V}^{J,\ell} \mathbf{S}_{\text{bf}}^{\ell-1} &= (\mathbf{U}_{2^{J-\ell+1}} \otimes \mathbf{I}_{2^{\ell-1}}) (\mathbf{I}_{2^{J-\ell+1}} \otimes \mathbf{U}_2 \otimes \mathbf{I}_{2^{\ell-2}}) \\ &\stackrel{(1.2)}{=} (\mathbf{U}_{2^{J-\ell+1}} \mathbf{I}_{2^{J-\ell+1}}) \otimes (\mathbf{I}_{2^{\ell-1}} (\mathbf{U}_2 \otimes \mathbf{I}_{2^{\ell-2}})) \\ &\stackrel{(1.2)}{=} \mathbf{U}_{2^{J-\ell+1}} \otimes \mathbf{U}_2 \otimes \mathbf{I}_{2^{\ell-2}} = \mathbf{U}_{2^{J-\ell}} \otimes \mathbf{I}_{2^{\ell-2}} = \mathbf{V}^{J,\ell-1}. \end{aligned}$$

Consequently, $\text{supp}(\mathbf{X}^{(J)} \dots \mathbf{X}^{(\ell)} \mathbf{X}^{(\ell-1)}) \subseteq \text{supp}(\mathbf{V}^{J,\ell} \mathbf{S}_{\text{bf}}^{\ell-1}) = \text{supp}(\mathbf{V}^{J,\ell-1}) = \mathbf{V}^{J,\ell-1}$. This ends the proof by induction. Now, in the case $q < J$, the butterfly support \mathbf{S}_{bf}^q of size $N \times N$ is block diagonal, where each block is simply the leftmost butterfly support of size $2^q \times 2^q$. Applying the previous case on each of these blocks of size $2^q \times 2^q$ yields the claimed result. ■

This sparsity structure of partial products will be exploited to establish identifiability for the Exact Matrix Factorization (EMF) - see (2.1) - of $\mathbf{H}^{(q,p)} := (\mathbf{X}^{(q)} \dots \mathbf{X}^{(\ell+1)}) (\mathbf{X}^{(\ell)} \dots \mathbf{X}^{(p)})$ (where $1 \leq p \leq \ell < q \leq J$) into two factors, when fixing the supports $\mathbf{W}^{(q:\ell+1)}$ and $\mathbf{W}^{(\ell:p)\top}$ as the respective sparsity constraint for the left and right factors. Indeed, this pair of supports have disjoint rank-one supports, so it satisfies the conditions of Proposition 2.7. While the role of the left factor \mathbf{X} in (2.1) will be played by $\mathbf{X}^{(q)} \dots \mathbf{X}^{(\ell+1)}$, that of the right factor \mathbf{Y}^\top will be played by $\mathbf{X}^{(\ell)} \dots \mathbf{X}^{(p)}$, so that the support constraint is indeed $\text{supp}(\mathbf{Y}) \subseteq \mathbf{W}^{(\ell:p)\top}$.

Lemma 3.7. Denoting $\mathbf{S}^L := \mathbf{W}^{(q:\ell+1)}$, and $\mathbf{S}^R := \mathbf{W}^{(\ell:p)\top}$, the N -tuple of rank-one supports $\varphi(\mathbf{S}^L, \mathbf{S}^R)$ has disjoint rank-one supports, for any $1 \leq p \leq \ell < q \leq J$. Consequently:

$$\mathcal{U}_s \left(\Sigma_{(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})} \right) = \left\{ (\mathbf{X}, \mathbf{Y}) \in \Sigma_{(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})} \mid \text{colsupp}(\mathbf{X}) = \text{colsupp}(\mathbf{Y}) = \llbracket N \rrbracket \right\}.$$

Proof. Denote $\mathcal{S} := \varphi(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})$. We also write $I_k := \{(k-1)2^\ell + 1, \dots, k2^\ell - 1\}$ for $k \in \llbracket \frac{N}{2^\ell} \rrbracket$. On the one hand, the right support $\mathbf{W}^{(\ell:p)\top}$, which is equal to $\mathbf{W}^{(\ell:p)}$ by Remark 3.5, is block diagonal, with blocks $\mathbf{V}^{\ell,p}$ of size $2^\ell \times 2^\ell$. This means that for $i \in I_k$ and $j \in I_{k'}$ with $k, k' \in \llbracket \frac{N}{2^\ell} \rrbracket$, $k \neq k'$, the rank-one supports \mathcal{S}^i and \mathcal{S}^j are disjoint. On the other hand, viewing column supports as subsets of indices, the columns of the left support $\mathbf{W}^{(q:\ell+1)}$ restricted to I_k are pairwise disjoint for each $k \in \llbracket \frac{N}{2^\ell} \rrbracket$, by definition of $\mathbf{V}^{q,\ell+1}$. This means that for $i, j \in I_k$ ($k \in \llbracket \frac{N}{2^\ell} \rrbracket$), the rank-one supports \mathcal{S}^i and \mathcal{S}^j are also disjoint, when $i \neq j$. Hence, by Proposition 2.7, we have $\mathcal{U}_s \left(\Sigma_{(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})} \right) = \text{IC}_\mathcal{S} \cap \text{MC}_\mathcal{S}$. We conclude by remarking that $(\mathbf{X}, \mathbf{Y}) \in \text{IC}_\mathcal{S} \cap \text{MC}_\mathcal{S}$ if, and only if, \mathbf{X} and \mathbf{Y} have no zero column. ■

Finally, we characterize the factors $\mathbf{X}^{(q)}, \dots, \mathbf{X}^{(\ell+1)}, \mathbf{X}^{(\ell)}, \dots, \mathbf{X}^{(p)}$ having a support included in the butterfly supports such that the partial product $\mathbf{X}^{(q)} \dots \mathbf{X}^{(\ell+1)}$ (resp. $\mathbf{X}^{(\ell)} \dots \mathbf{X}^{(p)}$) has no zero column (resp. no zero row). The proof is deferred to Appendix A.

Lemma 3.8. Let \mathbf{S}_{bf} be the butterfly supports of size $N = 2^J$, and consider $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{\mathbf{S}_{\text{bf}}}$. The following are equivalent:

- (i) for each $1 \leq p \leq \ell < q \leq J$, the partial product $\mathbf{X}^{(q)} \dots \mathbf{X}^{(\ell+1)}$ has no zero column, and the partial product $\mathbf{X}^{(\ell)} \dots \mathbf{X}^{(p)}$ has no zero row;
- (ii) $\mathbf{X}^{(J)}$ has no zero column, each factor $\mathbf{X}^{(\ell)}$ has no zero column and no zero row for $2 \leq \ell \leq J-1$, and $\mathbf{X}^{(1)}$ has no zero row.

3.2. Hierarchical factorization method. The hierarchical matrix factorization [12] is a method to recover the sparse factors $\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}$ from the product $\mathbf{Z} = \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$, by successively recovering the factors through several hierarchical levels. At the first level, the factors $\mathbf{X}^{(J)}$ and $\mathbf{H}^{(J-1)} := \mathbf{X}^{(J-1)} \dots \mathbf{X}^{(1)}$ are recovered from \mathbf{Z} . At the second level, the factors $\mathbf{X}^{(J-1)}$ and $\mathbf{H}^{(J-2)} := \mathbf{X}^{(J-2)} \dots \mathbf{X}^{(1)}$ are recovered from $\mathbf{H}^{(J-1)}$. The process is repeated recursively, until all the sparse factors are recovered. At each step of this hierarchical factorization, adequate support constraints are enforced on the factors: they correspond to sparsity patterns obtained from the partial product of several sparse factors.

When the sparse factorization at each hierarchical level is unique up to scaling ambiguities for these adequate support constraints, the recovered factors $\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}$ are shown to be the S-unique multi-layer exact matrix factorization (MEMF) of \mathbf{Z} for the considered constraints, in the sense of the following definition, which is a generalization of Definition 2.6.

Definition 3.9 (S-uniqueness of an MEMF in Σ). *For any set of J -tuples of factors $\Sigma \subseteq \mathbb{C}^{N_J \times N_{J-1}} \times \dots \times \mathbb{C}^{N_1 \times N_0}$ for some integers N_0, \dots, N_J , $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)})$ is the S-unique MEMF of $\mathbf{Z} := \mathbf{X}^{(J)} \mathbf{X}^{(J-1)} \dots \mathbf{X}^{(1)}$ in Σ , if, for any $(\mathbf{X}'^{(J)}, \dots, \mathbf{X}'^{(1)}) \in \Sigma$ such that $\mathbf{Z} := \mathbf{X}'^{(J)} \dots \mathbf{X}'^{(1)}$, there exist invertible diagonal matrices $\mathbf{D}^{(1)} \in \mathcal{D}_{N_1}, \dots, \mathbf{D}^{(J-1)} \in \mathcal{D}_{N_{J-1}}$ such that $\mathbf{X}'^{(\ell)} = \mathbf{D}^{(\ell-1)} \mathbf{X}^{(\ell)} \mathbf{D}^{(\ell)}$ for all $\ell \in \llbracket J \rrbracket$, with the convention $\mathbf{D}^{(J)} = \mathbf{I}_{N_J}$ and $\mathbf{D}^{(0)} = \mathbf{I}_{N_0}$.*

In our case, the factors are constrained to the butterfly supports, as introduced in the previous subsection. Based on the previous lemmas and on this hierarchical factorization method, we obtain the main result of this section, namely, the hierarchical identifiability in multi-layer sparse matrix factorization with butterfly structure.

Theorem 3.10. *Let \mathbf{S}_{bf} be the butterfly supports of size $N = 2^J$, and consider $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{\mathbf{S}_{\text{bf}}}$. Suppose that: $\mathbf{X}^{(J)}$ has no zero column, each $\mathbf{X}^{(\ell)}$ has no zero column and no zero row for $2 \leq \ell \leq J-1$, and $\mathbf{X}^{(1)}$ has no zero row. Then, the factors $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)})$ are the S-unique MEMF of $\mathbf{Z} := \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$ in $\Sigma_{\mathbf{S}_{\text{bf}}}$. Moreover, these factors can be recovered from \mathbf{Z} , up to scaling ambiguities only, through a hierarchical factorization method detailed in Algorithm 3.1 (further discussed below), where \mathbf{Z} and any partitioning binary tree \mathcal{T} of $\llbracket J \rrbracket$ (see Definition 3.12 below) are given as the inputs of the algorithm.*

This theorem can be applied to show identifiability of the butterfly factorization of the DFT matrix as suggested in [10, Chapter 7], but also the one of the Hadamard matrix.

Corollary 3.11. *The Hadamard matrix and the permuted DFT matrix $\mathbf{DFT}_N \mathbf{R}_N^\top$ of size N , where \mathbf{R}_N is the bit-reversal permutation matrix defined by (3.7), admit respectively an S-unique MEMF in $\Sigma_{\mathbf{S}_{\text{bf}}}$. In both cases, the butterfly factors can be recovered up to scaling ambiguities via Algorithm 3.1 with any partitioning binary tree of $\llbracket J \rrbracket$ as input.*

Before discussing the concrete details of Algorithm 3.1, let us mention that, in the specific case of the butterfly support constraint, the hierarchical factorization method previously described can be generalized to any kind of binary tree structure. For instance, in the case of four factors $\mathbf{X}^{(4)}, \dots, \mathbf{X}^{(1)}$ of size 16×16 , one can perform hierarchical factorization in an alternative fashion, by factorizing $\mathbf{Z} := \mathbf{X}^{(4)} \mathbf{X}^{(3)} \mathbf{X}^{(2)} \mathbf{X}^{(1)}$ into $\mathbf{X}^{(4)} \mathbf{X}^{(3)}$ and $\mathbf{X}^{(2)} \mathbf{X}^{(1)}$ at the first layer, then $\mathbf{X}^{(4)} \mathbf{X}^{(3)}$ into $\mathbf{X}^{(4)}, \mathbf{X}^{(3)}$, and finally $\mathbf{X}^{(2)} \mathbf{X}^{(1)}$ into $\mathbf{X}^{(2)}, \mathbf{X}^{(1)}$. Let us formally define such a tree structure that describes the factorization order in the hierarchical method.

Definition 3.12 (Partitioning binary tree). A partitioning binary tree of a $\{p, \dots, q\}$ is a binary tree, where nodes are non-empty subsets of $\{p, \dots, q\}$, which satisfies the axioms:

- (i) Each node is a subset of consecutive indices in $\{p, \dots, q\}$.
- (ii) The root is the set $\{p, \dots, q\}$.
- (iii) A node is a singleton if, and only if, it is a leaf.
- (iv) For each non-leaf node, the left and right children form a partition of their parent, in a such way that the indices of the left child are larger than those in the right child.

Examples of partitioning binary trees are illustrated in [Figure 3](#).

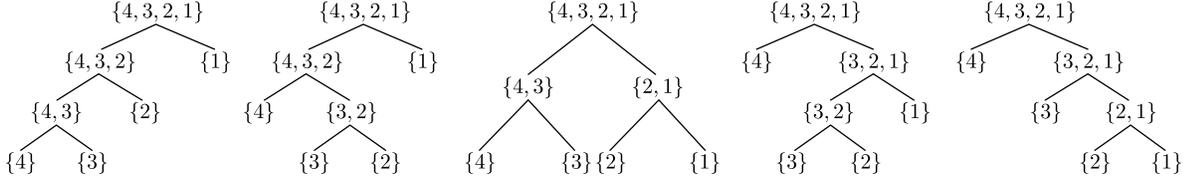


Figure 3: All the five possible partitioning binary trees of $\{1, 2, 3, 4\}$. Applying [Algorithm 3.1](#) with any of these trees and $\mathbf{Z} := \mathbf{X}^{(4)} \dots \mathbf{X}^{(1)}$ as inputs of the algorithm yields the exact recovery of the sparse factors $\mathbf{X}^{(4)}, \dots, \mathbf{X}^{(1)}$, provided that: $\mathbf{X}^{(4)}$ has no zero column; $\mathbf{X}^{(3)}, \mathbf{X}^{(2)}$ have no zero column and no zero row; and $\mathbf{X}^{(1)}$ has no zero row. Let us highlight here that this result is true for any number of layers J , and is not limited to the case $J = 4$.

Algorithm 3.1 Hierarchical butterfly factorization method of size $N = 2^J$.

Require: Integers $1 \leq p \leq q \leq J$, matrix $\mathbf{H}^{(q:p)} \in \mathbb{C}^{N \times N}$, partitioning binary tree $\mathcal{T}^{(q:p)}$ of $\{p, \dots, q\}$

- 1: **if** $q = p$ **then**
 - 2: **return** $\mathbf{H}^{(q:p)}$
 - 3: **end if**
 - 4: $\ell \leftarrow$ maximum value in the right child of the root of $\mathcal{T}^{(q:p)}$
 - 5: $(\mathcal{T}^{(q:\ell+1)}, \mathcal{T}^{(\ell:p)}) \leftarrow$ left and right subtrees of the root of $\mathcal{T}^{(q:p)}$
 - 6: $(\mathcal{S}^i)_{i=1}^N \leftarrow \varphi \left(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top} \right)$
 - 7: **for** $i = 1$ **to** N **do**
 - 8: $\mathcal{C}^i \leftarrow \text{BestRankOneApprox} \left((\mathbf{H}^{(q:p)})_{j,k} \right)_{(j,k) \in \mathcal{S}^i}$, where \mathcal{S}^i is viewed as an index set
 - 9: **end for**
 - 10: Set $(\mathbf{H}^{(q:\ell+1)}, \mathbf{H}^{(\ell:p)})$ as a pair of factors in $\varphi^{-1}(\{\mathcal{C}\}) \cap \text{IC}_{(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})}$
 - 11: $(\mathbf{X}'^{(q)}, \dots, \mathbf{X}'^{(\ell+1)}) \leftarrow$ result of [Algorithm 3.1](#) with inputs $\mathbf{H}^{(q:\ell+1)}$ and $\mathcal{T}^{(q:\ell+1)}$
 - 12: $(\mathbf{X}'^{(\ell)}, \dots, \mathbf{X}'^{(p)}) \leftarrow$ result of [Algorithm 3.1](#) with inputs $\mathbf{H}^{(\ell:p)}$ and $\mathcal{T}^{(\ell:p)}$
 - 13: **return** $(\mathbf{X}'^{(q)}, \dots, \mathbf{X}'^{(\ell+1)}, \mathbf{X}'^{(\ell)}, \dots, \mathbf{X}'^{(p)})$
-

We are now able to describe [Algorithm 3.1](#), given as inputs a partitioning binary tree \mathcal{T} of $\llbracket J \rrbracket$ and any target matrix \mathbf{Z} . The algorithm visits the nodes of \mathcal{T} in a breadth-first order, starting by the root node. At each non-leaf node $n \subseteq \llbracket J \rrbracket$ characterized by its maximum

value q , its minimum value p , and its “splitting index” ℓ , which is the maximum value of its right child, the algorithm performs an approximation of the intermediate matrix $\mathbf{H}^{(q:p)}$ by a sum of rank-one matrices $(\mathbf{C}^i)_{i=1}^N$, whose supports are constrained by the rank-one supports $(\mathbf{S}^i)_{i=1}^N = \varphi(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})$. This is done by defining \mathbf{C}^i ($1 \leq i \leq N$) as the best rank-one approximation (written `BestRankOneApprox` in [Algorithm 3.1](#)) of the submatrix $(\mathbf{H}^{(q:p)})_{j,k} \text{ on } (j,k) \in \mathcal{S}^i$, which can be computed via a truncated singular value decomposition (SVD). This yields an approximation of $\mathbf{H}^{(q:p)}$ by a product of two factors $\mathbf{H}^{(q:\ell+1)}\mathbf{H}^{(\ell:p)\top}$, where the left (resp. right) factor $\mathbf{H}^{(q:\ell+1)}$ (resp. $\mathbf{H}^{(\ell:p)\top}$) have a support included in $\mathbf{W}^{(q:\ell+1)}$ (resp. $\mathbf{W}^{(\ell:p)\top}$), and are optimal as proved in [9] in the sense that $\|\mathbf{H}^{(q:p)} - \mathbf{X}\mathbf{Y}^\top\|_F^2$ is minimized among all \mathbf{X}, \mathbf{Y} satisfying the same support constraints.

Assume now that \mathbf{Z} admits an exact butterfly factorization $\mathbf{Z} = \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$ where $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{\text{Sbf}}$ satisfies the conditions of [Theorem 3.10](#). Then, the proof for [Theorem 3.10](#) relies on two key steps. Firstly, we show recursively that the intermediate matrix to factorize at each node $n = \{p, \dots, q\}$ of \mathcal{T} in [Algorithm 3.1](#) is shaped as $\mathbf{H}^{(q:p)} = \mathbf{D}^{(q)-1}\mathbf{X}^{(q)} \dots \mathbf{X}^{(p)}\mathbf{D}^{(p-1)}$, where $\mathbf{D}^{(q)}, \mathbf{D}^{(p-1)}$ are invertible diagonal matrices. Secondly, we show that $\mathbf{H}^{(q:p)}$ admits an S-unique EMF in $\Sigma_{(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})}$, as claimed in the following.

Lemma 3.13. *Let Sbf be the butterfly supports of size 2^J , and consider any $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{\text{Sbf}}$. Suppose that: $\mathbf{X}^{(J)}$ has no zero column, each $\mathbf{X}^{(\ell)}$ has no zero column and no zero row for $2 \leq \ell \leq J-1$, and $\mathbf{X}^{(1)}$ has no zero row. Then, for any $1 \leq p \leq \ell < q \leq J$, for any invertible diagonal matrices $\mathbf{D}, \mathbf{D}' \in \mathcal{D}_N$, the matrix $\mathbf{D}^{-1}\mathbf{X}^{(q)} \dots \mathbf{X}^{(\ell+1)}\mathbf{X}^{(\ell)} \dots \mathbf{X}^{(p)}\mathbf{D}'$ admits an S-unique EMF in $\Sigma_{(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})}$, which is $(\mathbf{D}^{-1}\mathbf{X}^{(q)} \dots \mathbf{X}^{(\ell+1)}, (\mathbf{X}^{(\ell)} \dots \mathbf{X}^{(p)}\mathbf{D}')^\top)$.*

Proof. Denote $\mathbf{H}^{(q:\ell+1)} := \mathbf{X}^{(q)} \dots \mathbf{X}^{(\ell+1)}$, and $\mathbf{H}^{(\ell:p)} := (\mathbf{X}^{(\ell)} \dots \mathbf{X}^{(p)})^\top$. By [Lemma 3.4](#), $(\mathbf{H}^{(q:\ell+1)}, \mathbf{H}^{(\ell:p)}) \in \Sigma_{(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})}$. By [Lemma 3.8](#) and the assumption on the factors $\mathbf{X}^{(\ell)}$, $1 \leq \ell \leq J$, the matrices $\mathbf{H}^{(q:\ell+1)}$ and $\mathbf{H}^{(\ell:p)}$ have no zero column. The same is true for $\mathbf{D}^{-1}\mathbf{H}^{(q:\ell+1)}$ and $\mathbf{D}'\mathbf{H}^{(\ell:p)}$, as the multiplication of a matrix by \mathbf{D}^{-1} or \mathbf{D}' does not change its support. By [Lemma 3.7](#), we obtain $(\mathbf{D}^{-1}\mathbf{H}^{(q:\ell+1)}, \mathbf{D}'\mathbf{H}^{(\ell:p)}) \in \mathcal{U}_s(\Sigma_{(\mathbf{W}^{(q:\ell+1)}, \mathbf{W}^{(\ell:p)\top})})$. \blacksquare

Proof of Theorem 3.10. For each node $n \subseteq \{p, \dots, q\}$ in the partitioning binary tree \mathcal{T} , we write q_n and p_n respectively the maximum and minimum index of node n . When n has children, ℓ_n is the “splitting” index of node n , which is the maximum value of its right child.

Let $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}), (\mathbf{X}'^{(J)}, \dots, \mathbf{X}'^{(1)}) \in \Sigma_{\text{Sbf}}$ such that $\mathbf{X}'^{(J)} \dots \mathbf{X}'^{(1)} = \mathbf{Z} := \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$, and suppose that $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)})$ verifies the assumptions of [Theorem 3.10](#). For any $1 \leq p \leq q \leq J$, denote $\mathbf{H}^{(q:p)} := \mathbf{X}^{(q)} \dots \mathbf{X}^{(p)}$, and $\mathbf{H}'^{(q:p)} := \mathbf{X}'^{(q)} \dots \mathbf{X}'^{(p)}$. Consider the sequence of nodes (n_1, n_2, \dots, n_w) in \mathcal{T} that are obtained after a breadth-first search of the binary tree, where w is the total number of nodes in \mathcal{T} . Then, consider the subsequence of (n_1, n_2, \dots, n_w) where we keep only nodes that are not leaves. Denote them $n_{\gamma(1)}, \dots, n_{\gamma(v)}$, with v the total number of non-leaf nodes in \mathcal{T} , and γ being an increasing function. To simplify, we denote $\ell_{\gamma(u)} := \ell_{n_{\gamma(u)}}$, $q_{\gamma(u)} := q_{n_{\gamma(u)}}$, and $p_{\gamma(u)} := p_{n_{\gamma(u)}}$, for $u \in \llbracket v \rrbracket$. [Figure 4](#) illustrates these notations on an example.

For any $u \in \llbracket v \rrbracket$, the assertion P_u is defined as: “there exist invertible diagonal matrices $\mathbf{D}^{(\ell_{\gamma(1)})}, \dots, \mathbf{D}^{(\ell_{\gamma(u)})} \in \mathcal{D}_N$ such that, for each $1 \leq \eta \leq u$, we have $\mathbf{H}'^{(q_{\gamma(\eta)}:\ell_{\gamma(\eta)+1})} = \mathbf{D}^{(q_{\gamma(\eta)})^{-1}}\mathbf{H}^{(q_{\gamma(\eta)}:\ell_{\gamma(\eta)+1})}\mathbf{D}^{(\ell_{\gamma(\eta)})}$ and $\mathbf{H}'^{(\ell_{\gamma(\eta)}:p_{\gamma(\eta)})^\top} = \mathbf{D}^{(\ell_{\gamma(\eta)})^{-1}}\mathbf{H}^{(\ell_{\gamma(\eta)}:p_{\gamma(\eta)})^\top}\mathbf{D}^{(p_{\gamma(\eta)}-1)}$ ”, with

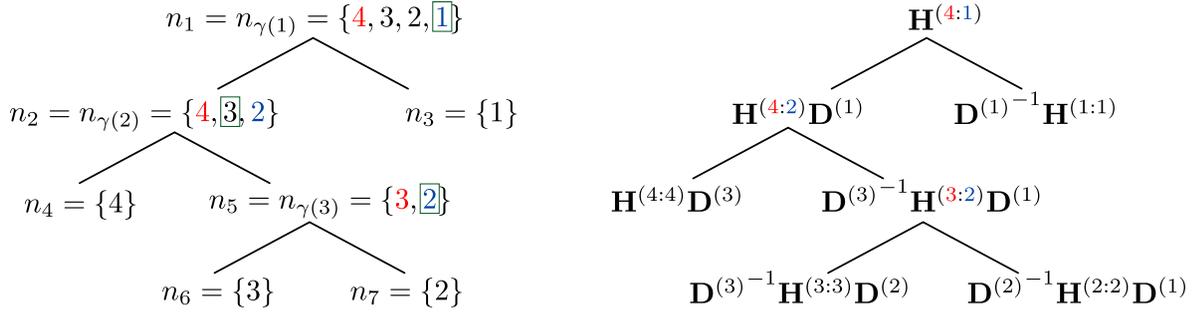


Figure 4: Illustration of the proof of [Theorem 3.10](#) on a given partitioning binary tree \mathcal{T} of $\{1, 2, 3, 4\}$. Left: Illustration of the breadth-first search order. We extract the non-leaf nodes using the increasing function γ . At each node n , the maximum value q_n is colored in red, the minimum value p_n is colored in blue, and the “splitting” index ℓ_n , which is the maximum of the right child of n when n is not a leaf, is boxed in green. Right: Illustration of the recursion in the proof of [Theorem 3.10](#).

the convention $\mathbf{D}^{(J)} = \mathbf{D}^{(0)} = \mathbf{I}_{\mathbf{N}}$. Let us show by induction that P_u is true for all $u \in \llbracket v \rrbracket$. By [Lemma 3.13](#), P_1 is true, because $n_{\gamma(1)}$ is the root node by definition. Let $u \in \{2, \dots, v\}$, and suppose that P_{u-1} is true. Fix $\mathbf{D}^{(\ell_{\gamma(1)})}, \dots, \mathbf{D}^{(\ell_{\gamma(u-1)})} \in \mathcal{D}_{\mathbf{N}}$ that verify P_{u-1} . By definition of the breadth-first search order, the parent of $n_{\gamma(u)}$ is necessarily a node $n_{\gamma(\eta)}$ with $1 \leq \eta \leq u-1$. Hence, either $n_{\gamma(u)} = \{\ell_{\gamma(\eta)} + 1, \dots, q_{\gamma(\eta)}\}$, or $n_{\gamma(u)} = \{p_{\gamma(\eta)}, \dots, \ell_{\gamma(\eta)}\}$, depending whether $n_{\gamma(u)}$ is a left or right child of $n_{\gamma(\eta)}$. Without loss of generality, we suppose the former, as the proof is similar if we suppose the latter. By assumption P_{u-1} , we have $\mathbf{H}'^{(q_{\gamma(u)}:p_{\gamma(u)})} = \mathbf{H}'^{(q_{\gamma(\eta)}:\ell_{\gamma(\eta)}+1)} = \mathbf{D}^{(q_{\gamma(\eta)})^{-1}}\mathbf{H}^{(q_{\gamma(\eta)}:\ell_{\gamma(\eta)}+1)}\mathbf{D}^{(\ell_{\gamma(\eta)})} = \mathbf{D}^{(q_{\gamma(u)})^{-1}}\mathbf{H}^{(q_{\gamma(u)}:p_{\gamma(u)})}\mathbf{D}^{(p_{\gamma(u)}-1)}$, because $p_{\gamma(u)} = \ell_{\gamma(\eta)} + 1$ and $q_{\gamma(u)} = q_{\gamma(\eta)}$. Hence:

$$\left(\mathbf{H}'^{(q_{\gamma(u)}:\ell_{\gamma(u)}+1)}\right)\left(\mathbf{H}'^{(\ell_{\gamma(u)}:p_{\gamma(u)})}\right) = \left(\mathbf{D}^{(q_{\gamma(u)})^{-1}}\mathbf{H}^{(q_{\gamma(u)}:\ell_{\gamma(u)}+1)}\right)\left(\mathbf{H}^{(\ell_{\gamma(u)}:p_{\gamma(u)}+1)}\mathbf{D}^{(\ell_{\gamma(u)}-1)}\right).$$

By [Lemma 3.13](#), P_u is verified. This is true for any $u \in \llbracket v \rrbracket$, and P_v yields our claim. \blacksquare

Remark 3.14. One can exploit [Algorithm 3.1](#) beyond the exact setting to approximate any matrix \mathbf{Z} of size 2^J by a matrix having the butterfly structure. Indeed, as proved in [\[9\]](#), the procedure at [Line 8](#) yields an optimal approximation (in the sense of the Frobenius norm) of each submatrix $(\mathbf{Z}_{k,l})_{(k,l) \in \mathcal{S}^i}$ as a product of two factors with the prescribed supports. However as this is used in a recursive greedy fashion in the algorithm, global optimality of the resulting multi-layer factorization is not necessarily guaranteed. Understanding the stability of the algorithm beyond exact recovery is an interesting challenge left to future work.

3.3. Complexity bounds. Existing algorithms for butterfly factorization [\[3, 12\]](#) are based on gradient descent, and as such they require to tune several criteria such as learning rate or stopping criteria. In contrast, [Algorithm 3.1](#) has a bounded complexity as it essentially consists in a controlled number of partial SVDs to compute rank-one approximations of submatrices. While full SVD of a matrix of size $m \times n$ would require $\mathcal{O}(mn \min(m, n))$ flops, partial SVD

with numerical rank k requires only $\mathcal{O}(kmn)$ flops (see e.g. [6] and references therein). Hence, in our complexity analysis of [Algorithm 3.1](#), the theoretical complexity of computing the best rank-one approximation of a matrix of size $m \times n$ will be $\mathcal{O}(mn)$. Below we estimate and compare the complexity for two types of partitioning binary trees.

Unbalanced tree. First we consider running [Algorithm 3.1](#) with a matrix \mathbf{Z} of size $N \times N$, $N = 2^J$ ($J \geq 2$), and the *unbalanced* partitioning binary tree \mathcal{T} of $\llbracket J \rrbracket$ (which is defined as the partitioning binary tree where the left child of each non-leaf node is a singleton) as inputs. There are in total $J - 1$ non-leaf nodes in this tree. At the non-leaf node of depth $j \in \{0, \dots, J - 2\}$, the algorithm computes the best rank-one approximation of N submatrices of size $2 \times N/2^{j+1}$, which yields a cost of the order of $N \times (2 \times N/2^{j+1}) = N^2/2^j$. Hence, the total cost of [Algorithm 3.1](#) with the unbalanced partitioning binary tree is of the order of:

$$\sum_{j=0}^{J-2} \frac{N^2}{2^j} = \mathcal{O}(2N^2).$$

Balanced tree. Let us now consider running [Algorithm 3.1](#) with a matrix \mathbf{Z} of size $N \times N$, $N = 2^J$ where J itself is also a power of 2, and the *balanced* partitioning binary tree \mathcal{T} of $\llbracket J \rrbracket$ (which is defined as the partitioning binary tree where the children of each non-leaf node have the same cardinality) as inputs. At each non-leaf node of depth $k \in \{0, \dots, \log_2(J) - 1\}$, the algorithm computes the best rank-one approximation of N submatrices of size $\sqrt{N^{1/2^k}} \times \sqrt{N^{1/2^k}}$, which yields a cost of the order of $N \times (\sqrt{N^{1/2^k}} \times \sqrt{N^{1/2^k}}) = N \times N^{1/2^k}$. In total, there are 2^k nodes of depth $k \in \{0, \dots, \log_2(J) - 1\}$. For $1 \leq k \leq \log_2(J) - 1$: $2^k \leq J/2 = \log_2(N)/2$. Hence, the total cost of [Algorithm 3.1](#) with the balanced tree is bounded as follow:

$$\sum_{k=0}^{\log_2(J)-1} 2^k N \times N^{1/2^k} = N^2 + \sum_{k=1}^{\log_2(J)-1} 2^k N^{1+1/2^k} \leq N^2 + \sum_{k=1}^{\log_2(\log_2(N))-1} \frac{\log_2(N)}{2} N^{3/2} = \mathcal{O}(N^2).$$

Discussion. Hence, the complexity of [Algorithm 3.1](#) is of the same order of magnitude as the one of matrix-vector multiplication of size $N \times N$, which is $\mathcal{O}(N^2)$. Suppose that we want to compute the product \mathbf{AB} , where \mathbf{A}, \mathbf{B} are of size $N \times N$. The naive computation requires $\mathcal{O}(N^3)$ flops. Assume now that \mathbf{A} admits a butterfly structure, i.e., there exists $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{\text{Sbf}}$ such that $\mathbf{A} = \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$. Then, one can compute such a butterfly factorization with [Algorithm 3.1](#), which requires only $\mathcal{O}(N^2)$ flops. This allows to compute $\mathbf{X}^{(J)} \dots \mathbf{X}^{(1)} \mathbf{B}$ with $\mathcal{O}(N^2 \log(N))$ flops, since the linear operator $\mathbf{b} \mapsto \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)} \mathbf{b}$ defined on \mathbb{C}^N is a fast transform which can be evaluated with only $\mathcal{O}(N \log N)$ flops [3]. In total, this method allows for a computation of \mathbf{AB} in $\mathcal{O}(N^2(\log N + 1))$ flops only, instead of $\mathcal{O}(N^3)$.

Finally, it is possible to implement [Algorithm 3.1](#) in a distributed fashion. Indeed, the computation of the best rank-one approximation of each submatrix at [Line 8](#) can be performed in parallel: in the setting with T threads, each thread computes the best rank-one approximation of $\lfloor N/T \rfloor$ submatrices. Moreover, when running [Algorithm 3.1](#) with a balanced partitioning binary tree, the implementation can be further parallelized, since the factorization at each node of the same depth can be performed by independent threads.

4. Identifiability in a family of sparsity patterns. We now investigate identifiability in multi-layer structured/sparse factorization *beyond the case of fixed supports*. We first start by

presenting such identifiability results in the two-layer setting for special matrices, including the DFT, DCT and DST matrices, and highlight that identifiability may dramatically fail for some other special cases, like the Hadamard matrix. Then, in the multi-layer setting, we show a stronger hierarchical identifiability result than [Corollary 3.11](#) for the butterfly factorization of the DFT matrix, by relaxing the sparsity constraint to a family of sparsity patterns.

4.1. Two-layer identifiability for DFT, DCT and DST matrices. The two-layer identifiability results are established for the DFT, DCT (type II) and DST (type II) matrices:

$$(4.1) \quad \mathbf{DCT}_N := \left(\cos \left[\frac{\pi}{N} \left(l - \frac{1}{2} \right) (k - 1) \right] \right)_{k,l \in \llbracket N \rrbracket},$$

$$(4.2) \quad \mathbf{DST}_N := \left(\sin \left[\frac{\pi}{N} \left(l - \frac{1}{2} \right) k \right] \right)_{k,l \in \llbracket N \rrbracket}.$$

Consider the family $\Lambda_{k\text{-col}}^{m \times n}$ of supports of size $m \times n$ which are k -sparse by column (the superscript is omitted when the matrix size is not ambiguous):

$$(4.3) \quad \Lambda_{k\text{-col}}^{m \times n} := \{ \mathbf{S} \in \mathbb{B}^{m \times n} \mid \forall j \in \llbracket n \rrbracket, \|\mathbf{S}_j\|_0 \leq k \}.$$

To establish our identifiability results in a family of sparsity patterns, we rely on [Theorem 2.4](#), which states that PS-uniqueness of an EMF in Σ_Ω for a given family of sparsity patterns Ω is equivalent to P-uniqueness of an Exact Matrix Decomposition (EMD) in Γ_Ω (see [Definition 2.3](#)), except for some trivial cases. We recall that the notations Γ_Ω , $\Gamma_{\mathcal{S}}$, $\mathcal{A}(\cdot)$ and $\mathcal{U}(\Gamma)$ have been introduced by [\(2.7\)](#), [\(2.8\)](#), [\(2.9\)](#), [\(2.10\)](#). Then, to characterize P-uniqueness of an EMD, the companion paper [\[20\]](#) suggests to employ the following strategy:

Proposition 4.1 ([\[20, Proposition 4.9\]](#)). *For any family of pairs of supports Ω stable by permutation, we have $\mathcal{C} \in \mathcal{U}(\Gamma_\Omega)$ if, and only if, both of the following conditions are verified:*

- (i) *For all $\mathcal{S} := \{\mathcal{S}^i\}_{i=1}^r \in \varphi(\Omega)$ such that $\mathcal{A}(\mathcal{C}) \in \mathcal{A}(\Gamma_{\mathcal{S}})$, there exists a permutation σ of $\llbracket r \rrbracket$ such that: $\text{supp}(\mathcal{C}^i) \subseteq \mathcal{S}^{\sigma(i)}$ for each $1 \leq i \leq r$.*
- (ii) *For all $\mathcal{S} \in \varphi(\Omega)$ such that $\mathcal{C} \in \Gamma_{\mathcal{S}}$, we have $\mathcal{C} \in \mathcal{U}(\Gamma_{\mathcal{S}})$.*

A simple sufficient condition for (i) can be established [\[20\]](#). To state it we denote

$$(4.4) \quad \varphi^{\text{cp}}(\Omega) := \bigcup_{\mathcal{S} \in \varphi(\Omega)} \left\{ \mathcal{S}' \in (\mathbb{B}^{m \times n})^r \mid \forall i = 1, \dots, r, \text{rank}(\mathcal{S}'^i) \leq 1, \mathcal{S}'^i \subseteq \mathcal{S}^i \right\}$$

and recall that $\mathcal{S}'^i \subseteq \mathcal{S}^i$ means the inclusion of the supports viewed as subsets of indices.

Proposition 4.2 ([\[20, Proposition 4.11\]](#)). *Consider $\mathcal{C} \in \Gamma_\Omega$, $\mathbf{Z} := \mathcal{A}(\mathcal{C})$, and assume that:*

- (i) *for each $\mathcal{C}' \in \Gamma_\Omega$ such that $\mathcal{A}(\mathcal{C}') = \mathbf{Z}$, the supports $\{\text{supp}(\mathcal{C}'^i)\}_{i=1}^r$ are pairwise disjoint;*
- (ii) *all $\mathcal{S} \in \varphi^{\text{cp}}(\Omega)$ such that $\{\mathcal{S}^i\}_{i=1}^r$ is a partition of $\text{supp}(\mathbf{Z})$, and such that the rank of $(\mathbf{Z}_{k,l})_{(k,l) \in \mathcal{S}^i}$ ($1 \leq i \leq r$) is at most one, are equivalent up to a permutation.*

Then, the supports $\{\text{supp}(\mathcal{C}^i)\}_{i=1}^r$ are identifiable in the sense of (i) in [Proposition 4.1](#).

We start by showing that the sufficient conditions given by [Proposition 4.2](#) to identify the rank-one supports constraint among a well-chosen family of sparsity patterns are verified for the exact matrix decomposition of the DFT, DCT and DST matrix.

Proposition 4.3. *Let $N = 2^J$ ($J \geq 2$). Using the notations introduced in (4.3), define $\Omega := \Lambda_{N/2-\text{col}} \times \Lambda_{2-\text{col}}$ the family of pairs of left and right supports of size $N \times N$ which are respectively $N/2$ -sparse by column and 2-sparse by column. Let \mathbf{Z} be either \mathbf{DFT}_N , \mathbf{DCT}_N , or \mathbf{DST}_N , and consider a tuple of rank-one matrices $\mathcal{C} \in \Gamma_\Omega$ such that $\mathbf{Z} = \mathcal{A}(\mathcal{C})$. Then, the sufficient conditions of Proposition 4.2 on \mathcal{C} and Ω are verified.*

Remark 4.4. We conjecture that similar results for the DCT and DST of type I and III can be established, but not for type IV. For the sake of conciseness, we only give identifiability results for the DCT-II and DST-II matrices with their formal proof, as type II is the most common type for DCT and DST. The extension for DCT and DST matrices of type I and III is therefore left as future work.

Informal proof of Proposition 4.3. Let us start by giving an informal proof for the case $\mathbf{Z} = \mathbf{DFT}_4$, as it illustrates concretely the assumptions of Proposition 4.2. Denoting $(\mathbf{S}_{\text{bf}}^2, \mathbf{S}_{\text{bf}}^1)$ as the butterfly supports of size 4, $\varphi(\mathbf{S}_{\text{bf}}^2, \mathbf{S}_{\text{bf}}^1)$ is only partition of $\text{supp}(\mathbf{DFT}_4)$ into four rank-one supports $(\mathcal{S}^1, \dots, \mathcal{S}^4)$, which have at most 2 nonzero rows and 2 nonzero columns, and such that the submatrix of \mathbf{DFT}_4 restricted to \mathcal{S}^i ($1 \leq i \leq 4$) is of rank one. We illustrate the submatrices restricted to these rank-one supports by representing the entries of the same submatrix with the same color: $\mathbf{DFT}_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$. ■

The formal proof of Proposition 4.3 for $\mathbf{Z} = \mathbf{DFT}_N$ is provided in [10, Chapter 7, Section 7.4]. The proof for $\mathbf{Z} = \mathbf{DCT}_N$ or $\mathbf{Z} = \mathbf{DST}_N$ is deferred to Appendix B. The main idea of the proof is that for any $\mathcal{C} \in \Gamma_\Omega$ such that $\mathbf{DCT}_N = \mathcal{A}(\mathcal{C})$ or $\mathbf{DST}_N = \mathcal{A}(\mathcal{C})$, each rank-one support $\text{supp}(\mathcal{C}^i)$, $1 \leq i \leq N$, corresponds to a set of $N/2$ rows with indices of the same parity, and of two columns with indices $\{l_1^i, l_2^i\}$ such that $(l_1^i + l_2^i - 1) = N$. An illustration of these supports is shown in Figure 5.

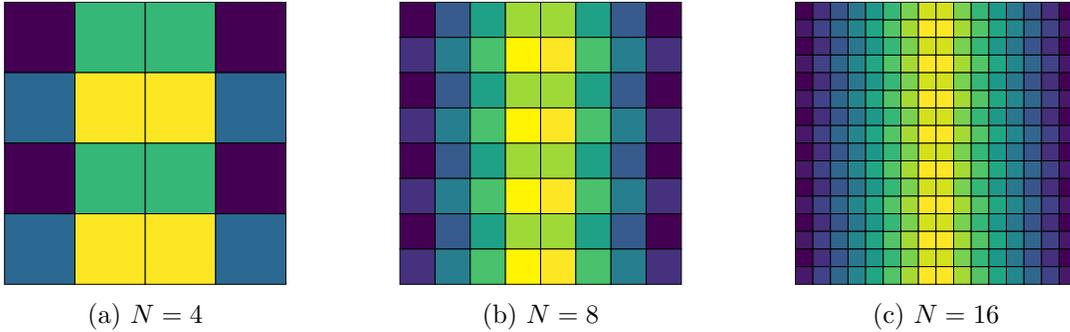


Figure 5: Illustration of the supports $\{\text{supp}(\mathcal{C}^i)\}_{i=1}^N$, where \mathcal{C} is an EMD of \mathbf{DCT}_N or \mathbf{DST}_N in Γ_Ω , with $\Omega := \Lambda_{N/2-\text{col}} \times \Lambda_{2-\text{col}}$, where $\Lambda_{k-\text{col}}$ has been defined in (4.3). Each support $\text{supp}(\mathcal{C}^i)$, $i \in \llbracket N \rrbracket$, is associated to a distinct color and all indices belonging to this support bear this color.

Equipped with Proposition 4.3 we are now able to show the following theorem. Hence,

our work generalizes the arguments of [10, Chapter 7] to prove that the sparse factorization of the DFT matrix into two factors is identifiable for the family $\Lambda_{N/2-\text{col}} \times \Lambda_{2-\text{col}}$.

Theorem 4.5. *Let \mathbf{Z} be either the DFT matrix, the DCT matrix or the DST matrix (type II) of size $N = 2^J$ ($J \geq 2$). Then, \mathbf{Z} has a PS-unique EMF in Σ_Ω where $\Omega := \Lambda_{N/2-\text{col}} \times \Lambda_{2-\text{col}}$. Moreover, \mathbf{DFT}_N also admits a PS-unique EMF in $\Sigma_{\Omega'}$ where $\Omega' := \Lambda_{2-\text{col}} \times \Lambda_{N/2-\text{col}}$.*

Proof. Suppose there exists $(\mathbf{X}, \mathbf{Y}) \in \Sigma_\Omega$ such that $\mathbf{Z} = \mathbf{X}\mathbf{Y}^\top$ and denote $\mathcal{C} := \varphi(\mathbf{X}, \mathbf{Y})$. We have $\mathcal{C} \in \Gamma_\Omega$, with naturally $\mathbf{Z} = \mathcal{A}(\mathcal{C})$. We verify that the conditions of Proposition 4.1 are satisfied in order to show $\mathcal{C} \in \mathcal{U}(\Gamma_\Omega)$. By Proposition 4.3, since $\mathcal{C} \in \Gamma_\Omega$ and $\mathbf{Z} = \mathcal{A}(\mathcal{C})$, the sufficient conditions of Proposition 4.2 are satisfied, hence the supports $\{\text{supp}(\mathbf{C}^i)\}_{i=1}^r$ are pairwise disjoint, and they are identifiable in the sense of condition (i) in Proposition 4.1. Let us show that condition (ii) of Proposition 4.1, i.e., fixed-support identifiability, is also verified. Consider a tuple of rank-one supports $\mathcal{S} \in \varphi(\Omega)$ such that $\mathcal{C} \in \Gamma_{\mathcal{S}}$.

First, we observe that the cardinality of $\text{supp}(\mathbf{C}^i)$ is equal to $\|\mathbf{C}^i\|_0 = N$ for all $i \in \llbracket N \rrbracket$: otherwise, since $\|\mathbf{C}^i\|_0 \leq N$ for $1 \leq i \leq N$, we would have $\|\mathbf{Z}\|_0 = \|\mathcal{A}(\mathcal{C})\|_0 = \|\sum_{j=1}^N \mathbf{C}^j\|_0 \leq \sum_{j=1}^N \|\mathbf{C}^j\|_0 < N^2 = \|\mathbf{Z}\|_0$. Since $\mathcal{C} \in \Gamma_{\mathcal{S}}$, for each i we have $\text{supp}(\mathbf{C}^i) \subseteq \mathcal{S}^i$, hence $N = \|\mathbf{C}^i\|_0 \leq \|\mathcal{S}^i\|_0 \leq N$ for all i , where the last inequality comes from the definition of the family Ω . It follows that $\text{supp}(\mathbf{C}^i) = \mathcal{S}^i$ for all i , i.e., $\mathcal{C} \in \Gamma_{\mathcal{S}}^{(=)}$, meaning that $\{\mathcal{S}^i\}_{i=1}^r$ are pairwise disjoint, since we showed above that $\{\text{supp}(\mathbf{C}^i)\}_{i=1}^r$ are pairwise disjoint. Hence, by Lemma 2.5, uniform P-uniqueness of EMD in $\Gamma_{\mathcal{S}}$ holds, i.e., $\mathcal{U}(\Gamma_{\mathcal{S}}) = \Gamma_{\mathcal{S}}$. As indeed $\mathcal{C} \in \Gamma_{\mathcal{S}}$ we obtain $\mathcal{C} \in \mathcal{U}(\Gamma_{\mathcal{S}})$. This shows as claimed that condition (ii) of Proposition 4.1 is satisfied. Therefore, by Proposition 4.1, $\varphi(\mathbf{X}, \mathbf{Y}) = \mathcal{C} \in \mathcal{U}(\Gamma_\Omega)$.

According to Theorem 2.4, in order to conclude that $(\mathbf{X}, \mathbf{Y}) \in \mathcal{U}(\Sigma_\Omega)$, there only remains to show $(\mathbf{X}, \mathbf{Y}) \in \text{IC}_\Omega \cap \text{MC}_\Omega$. If $(\mathbf{X}, \mathbf{Y}) \notin \text{IC}_\Omega$, there would exist i such that $\mathbf{C}^i = 0$, which would lead to a contradiction. Hence, $(\mathbf{X}, \mathbf{Y}) \in \text{IC}_\Omega$. We now show $(\mathbf{X}, \mathbf{Y}) \in \text{MC}_\Omega$. Consider a pair of supports $\mathcal{S} \in \Omega$ such that $(\mathbf{X}, \mathbf{Y}) \in \Sigma_{\mathcal{S}}$, and denote $\mathcal{S} := \varphi(\mathcal{S})$ the corresponding tuple of rank-one supports. Then, $\mathcal{C} \in \Gamma_{\mathcal{S}}$, and by repeating the previous argument, $\mathcal{C} \in \Gamma_{\mathcal{S}}^{(=)}$. This yields $\text{colsupp}(\mathbf{X}) = \text{colsupp}(\mathbf{S}^L)$ and $\text{colsupp}(\mathbf{Y}) = \text{colsupp}(\mathbf{S}^R)$, hence $(\mathbf{X}, \mathbf{Y}) \in \text{MC}_\Omega$ as claimed. By Theorem 2.4, we conclude that $(\mathbf{X}, \mathbf{Y}) \in \mathcal{U}(\Sigma_\Omega)$.

Finally, \mathbf{DFT}_N also admits a PS-unique EMF in $\Lambda_{2-\text{col}} \times \Lambda_{N/2-\text{col}}$ because $\mathbf{DFT}_N^\top = \mathbf{DFT}_N$ by its definition in (3.3). ■

4.2. A matrix for which identifiability fails. In [7, 8], a counter-example showed that the Hadamard matrix of size $N = 8$ does not admit a unique matrix factorization into two factors with $\Omega := \Lambda_{N/2-\text{col}} \times \Lambda_{2-\text{col}}$ as the family of allowed supports. They even show that one of these sparse factorization yields a factor that cannot be further factorized into sparse factors with a prescribed support pattern given by the family $\Lambda_{2-\text{col}} \times \Lambda_{2-\text{col}}$, which gives a pitfall of the hierarchical factorization method given the support constraint Ω at the first level of the hierarchy. Here, we generalize the non-uniqueness of the sparse factorization of the Hadamard matrix of size 2^J for any $J \geq 2$, given the family of sparsity patterns $\Omega := \Lambda_{N/2-\text{col}} \times \Lambda_{2-\text{col}}$.

Proposition 4.6. *Consider the Hadamard matrices \mathbf{H}_{2^J} ($J \geq 2$) defined as:*

$$(4.5) \quad \mathbf{H}_{2^J} := \underbrace{\mathbf{H}_2 \otimes \mathbf{H}_2 \otimes \dots \otimes \mathbf{H}_2}_{J \text{ times}} \in \mathbb{B}^{2^J \times 2^J}, \quad \text{where } \mathbf{H}_2 := \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

For each $J \geq 2$, \mathbf{H}_{2^J} admits an exact matrix factorization in Σ_Ω where $\Omega := \Lambda_{N/2-\text{col}} \times \Lambda_{2-\text{col}}$ and $N = 2^J$, but this factorization is not PS-unique.

The main idea of the proof, for which the details are deferred to [Appendix C](#), is to show that the equality $\mathbf{H}_2 \otimes \mathbf{H}_{2^{J-1}} = \mathbf{H}_{2^{J-1}} \otimes \mathbf{H}_2$ yields two non-equivalent factorizations of \mathbf{H}_{2^J} with supports in the prescribed Ω . The non-uniqueness of the sparse factorization of the Hadamard matrix may be related to the peculiar structure of this matrix, and we conjecture that [Theorem 4.5](#) actually holds for all matrices \mathbf{Z} except in some semi-algebraic set (of Lebesgue measure zero) of degenerate matrices. Further work is needed to confirm this result.

4.3. Hierarchical identifiability for the DFT matrix. Going back to the multi-layer case, we now show an improved hierarchical identifiability result for the butterfly factorization of the DFT matrix compared to [Corollary 3.11](#): we consider the setting in which the sparsity constraint is encoded by a family of sparsity patterns that relaxes the specific butterfly structure, in the sense that this family only enforces the block-diagonal structure constraints and the number of nonzero entries per column. Denote $\Lambda_{s-\text{bd}}^{n \times n}$ the family of supports of size $n \times n$ that are block-diagonal, with blocks of size $s \times s$ (we suppose that s divides n):

$$(4.6) \quad \Lambda_{s-\text{bd}}^{n \times n} := \left\{ \left(\begin{array}{ccc} \mathbf{S}^{(1)} & & 0 \\ & \ddots & \\ 0 & & \mathbf{S}^{(n/s)} \end{array} \right) \mid \mathbf{S}^{(1)}, \dots, \mathbf{S}^{(n/s)} \in \mathbb{B}^{s \times s} \right\}.$$

The superscript is omitted when there is no ambiguity. According to [\(3.6\)](#), the DFT matrix of size $N = 2^J$ admits a butterfly factorization, in the sense that $\mathbf{DFT}_N = \mathbf{F}^{(J)} \dots \mathbf{F}^{(1)} \mathbf{R}_N$, with $\mathbf{F}^{(\ell)}$ ($1 \leq \ell \leq J$) and \mathbf{R}_N defined by [\(3.6\)](#)-[\(3.7\)](#). Hence, the permuted DFT matrix $\mathbf{DFT}_N \mathbf{R}_N^\top$ admits a sparse factorization where the factors satisfy the following relaxed sparsity constraint:

$$(4.7) \quad \forall 1 \leq \ell \leq J, \quad \text{supp}(\mathbf{F}^{(\ell)}) = \mathbf{S}_{\text{bf}}^\ell \in \Lambda_{2-\text{col}} \cap \Lambda_{2^\ell-\text{bd}},$$

recalling that $(\mathbf{S}_{\text{bf}}^\ell)_{\ell=1}^J$ are the butterfly supports defined at [\(3.1\)](#). The rest of the subsection will show that this factorization is the unique one (up to scaling and permutation ambiguities), when the constraint set on the ℓ -th factor is encoded by the family $\Lambda_{2-\text{col}} \times \Lambda_{2^\ell-\text{bd}}$. For that purpose, one key remark is that the partial product $\mathbf{F}^{(q)} \dots \mathbf{F}^{(1)}$ is, up to some permutation of columns, equal to $\mathbf{I}_{N/2^q} \otimes \mathbf{DFT}_{2^q}$.

Lemma 4.7. *Consider $N = 2^J$ and $2 \leq q \leq J$. Then, with $\mathbf{F}^{(\ell)}, \mathbf{Q}^{(\ell)}$ ($1 \leq \ell \leq J$) defined as in [\(3.6\)](#)-[\(3.7\)](#), we have:*

$$(4.8) \quad \mathbf{F}^{(q)} \mathbf{F}^{(q-1)} \dots \mathbf{F}^{(1)} = (\mathbf{I}_{N/2^q} \otimes \mathbf{DFT}_{2^q}) \left(\mathbf{Q}^{(1)} \dots \mathbf{Q}^{(q)} \right)^\top.$$

Remark 4.8. This recursive block structure is specific to the DFT matrix and does not appear in the DCT or DST matrices, which explains why we only state our improved hierarchical identifiability result for the DFT matrix.

Proof. For $1 \leq \ell \leq J$, we have $\mathbf{F}^{(\ell)} = \mathbf{I}_{N/2^\ell} \otimes \mathbf{B}_{2^\ell}$ and $\mathbf{Q}^{(\ell)} = \mathbf{I}_{N/2^\ell} \otimes \mathbf{P}_{2^\ell}$ by definition, cf. (3.6)-(3.7). Hence $\mathbf{F}^{(J)} = \mathbf{B}_N$ and $\mathbf{Q}^{(J)} = \mathbf{P}_N$, and recursively applying (3.4) yields

$$\begin{aligned}
 \mathbf{DFT}_N &\stackrel{(3.4)}{=} \mathbf{B}_N [\mathbf{I}_2 \otimes \mathbf{DFT}_{N/2}] \mathbf{P}_N = \mathbf{F}^{(J)} [\mathbf{I}_2 \otimes \mathbf{DFT}_{N/2}] \mathbf{Q}^{(J)} \\
 &\stackrel{(3.4)}{=} \mathbf{F}^{(J)} [\mathbf{I}_2 \otimes (\mathbf{B}_{N/2} (\mathbf{I}_2 \otimes \mathbf{DFT}_{N/4}) \mathbf{P}_{N/2})] \mathbf{Q}^{(J)} \\
 &\stackrel{(1.2)}{=} \mathbf{F}^{(J)} [\mathbf{I}_2 \otimes \mathbf{B}_{N/2}] [\mathbf{I}_2 \otimes ((\mathbf{I}_2 \otimes \mathbf{DFT}_{N/4}) \mathbf{P}_{N/2})] \mathbf{Q}^{(J)} \\
 &\stackrel{(3.6)}{=} \mathbf{F}^{(J)} \mathbf{F}^{(J-1)} [\mathbf{I}_2 \otimes ((\mathbf{I}_2 \otimes \mathbf{DFT}_{N/4}) \mathbf{P}_{N/2})] \mathbf{Q}^{(J)} \\
 &\stackrel{(1.2)}{=} \mathbf{F}^{(J)} \mathbf{F}^{(J-1)} [\mathbf{I}_2 \otimes (\mathbf{I}_2 \otimes \mathbf{DFT}_{N/4})] [\mathbf{I}_2 \otimes \mathbf{P}_{N/2}] \mathbf{Q}^{(J)} \\
 &\stackrel{(3.7)}{=} \mathbf{F}^{(J)} \mathbf{F}^{(J-1)} [\mathbf{I}_4 \otimes \mathbf{DFT}_{N/4}] \mathbf{Q}^{(J-1)} \mathbf{Q}^{(J)} \\
 &= \dots \\
 &= \mathbf{F}^{(J)} \dots \mathbf{F}^{(q+1)} (\mathbf{I}_{N/2^q} \otimes \mathbf{DFT}_{2^q}) \mathbf{Q}^{(q+1)} \dots \mathbf{Q}^{(J)}.
 \end{aligned}$$

Moreover, by (3.6), we have $\mathbf{DFT}_N = \mathbf{F}^{(J)} \dots \mathbf{F}^{(q+1)} \mathbf{F}^{(q)} \dots \mathbf{F}^{(1)} \mathbf{R}_N$. Hence, $\mathbf{F}^{(q)} \dots \mathbf{F}^{(1)} = (\mathbf{I}_{N/2^q} \otimes \mathbf{DFT}_{2^q}) \mathbf{Q}^{(q+1)} \dots \mathbf{Q}^{(J)} \mathbf{R}_N^\top$ because $\mathbf{F}^{(J)}, \dots, \mathbf{F}^{(q+1)}$ are unitary as well as \mathbf{R}_N . This yields (4.8) by definition of \mathbf{R}_N , cf. (3.7), and unitarity of $\mathbf{Q}^{(\ell)}$. \blacksquare

As a corollary of Theorem 4.5, $\mathbf{I}_{N/2^q} \otimes \mathbf{DFT}_{2^q}$ admits a PS-unique EMF into two factors in a constraint set which enforces a block-diagonal structure and sparsity by column:

Corollary 4.9 (Application of Theorem 4.5). *Consider $N = 2^J$ and $2 \leq q \leq J$. Then, the block-diagonal matrix $\mathbf{I}_{N/2^q} \otimes \mathbf{DFT}_{2^q}$ admits a PS-unique EMF in Σ_{Ω_q} , where we define:*

$$(4.9) \quad \Omega_q := (\Lambda_{2-\text{col}} \cap \Lambda_{2^q-\text{bd}}) \times (\Lambda_{2^{q-1}-\text{col}} \cap \Lambda_{2^q-\text{bd}}).$$

Proof. By Theorem 4.5, the matrix \mathbf{DFT}_{2^q} admits a PS-unique EMF in $\Sigma_{\Lambda_{2-\text{col}} \times \Lambda_{2^{q-1}-\text{col}}}$. But for any $(\mathbf{S}^L, \mathbf{S}^R) \in \Omega_q$, the support \mathbf{S}^L (resp. \mathbf{S}^R) is block-diagonal where each block is of size $2^q \times 2^q$ and 2-sparse by column (resp. 2^{q-1} -sparse by column). Hence, by the block-diagonal structure of the family of sparsity patterns Ω_q , the block-diagonal matrix $\mathbf{I}_{N/2^q} \otimes \mathbf{DFT}_{2^q}$ admits a PS-unique EMF in Σ_{Ω_q} . \blacksquare

Consequently, we show that $\mathbf{F}^{(q)} \dots \mathbf{F}^{(1)}$ also admits a PS-unique EMF in Σ_{Ω_q} where Ω_q is defined by (4.9). This is essentially because the permutation matrix $\mathbf{Q} := \mathbf{Q}^{(1)} \dots \mathbf{Q}^{(q)}$ from the previous lemma is a permutation matrix that also has a block-diagonal structure.

Lemma 4.10. *Consider $N = 2^J$ and $2 \leq q \leq J$. Then, with $\mathbf{F}^{(\ell)}$ ($1 \leq \ell \leq J$) defined as in (3.6), the partial product $\mathbf{F}^{(q)} \dots \mathbf{F}^{(1)}$ admits a PS-unique EMF in the constraint set Σ_{Ω_q} where Ω_q is defined by (4.9), which is $(\mathbf{F}^{(q)}, (\mathbf{F}^{(q-1)} \dots \mathbf{F}^{(1)})^\top)$.*

Proof. Denote $\mathbf{X} := \mathbf{F}^{(q)}$, $\mathbf{Y} := (\mathbf{F}^{(q-1)} \dots \mathbf{F}^{(1)})^\top$, and $\mathbf{Z} := \mathbf{X}\mathbf{Y}^\top$. To prove $(\mathbf{X}, \mathbf{Y}) \in \mathcal{U}(\Sigma_{\Omega_q})$, consider $(\mathbf{X}', \mathbf{Y}') \in \Sigma_{\Omega_q}$ such that $\mathbf{X}'\mathbf{Y}'^\top = \mathbf{Z}$, and let us show $(\mathbf{X}', \mathbf{Y}') \sim (\mathbf{X}, \mathbf{Y})$. By (4.8), denoting $\mathbf{Q} := \mathbf{Q}^{(1)} \dots \mathbf{Q}^{(q)}$, we have $\mathbf{X}'\mathbf{Y}'^\top \mathbf{Q} = \mathbf{X}\mathbf{Y}^\top \mathbf{Q} = \mathbf{I}_{N/2^q} \otimes \mathbf{DFT}_{2^q}$. Denote $I_k = \{(k-1)2^q + 1, \dots, k2^q\}$ for any $k \in \llbracket \frac{N}{2^q} \rrbracket$. For any $1 \leq \ell \leq q$, the left multiplication of \mathbf{Y}' by $\mathbf{Q}^{(\ell)\top}$ cannot permute a row indexed by $i \in I_k$ with a row indexed by $j \in I_{k'}$ for $k \neq k'$, because

of the block structure of $\mathbf{Q}^{(\ell)}$ defined by (3.7). Hence, $\mathbf{Q}^\top \mathbf{Y}'$ is block diagonal where each block is of size $2^q \times 2^q$. Moreover, it is 2^{q-1} -sparse by column. In other words, $(\mathbf{X}', \mathbf{Q}^\top \mathbf{Y}') \in \Sigma_{\Omega_q}$. But by Corollary 4.9, $(\mathbf{X}, \mathbf{Q}^\top \mathbf{Y}) \in \mathcal{U}(\Sigma_{\Omega_q})$. Therefore, $(\mathbf{X}', \mathbf{Q}^\top \mathbf{Y}') \sim (\mathbf{X}, \mathbf{Q}^\top \mathbf{Y})$. This yields $(\mathbf{X}', \mathbf{Y}') \sim (\mathbf{X}, \mathbf{Y})$ as claimed, since \mathbf{Q} is unitary. \blacksquare

Consider now the family of sparsity patterns $\Omega'_q \subseteq \Omega_q$ which yields a slightly more restrictive sparsity constraint than the one encoded by Ω_q :

$$(4.10) \quad \Omega'_q := (\Lambda_{2-\text{col}} \cap \Lambda_{2^q-\text{bd}}) \times \Lambda_{2^{q-1}-\text{bd}} \subseteq \Omega_q, \quad \text{for } 2 \leq q \leq J.$$

As a consequence of Lemma 4.10, the partial product $\mathbf{F}^{(q)} \dots \mathbf{F}^{(1)}$ admits a PS-unique EMF in the constraint set $\Sigma_{\Omega'_q}$, simply because $\Sigma_{\Omega'_q} \subseteq \Sigma_{\Omega_q}$. However, this restriction allows us to establish a stronger uniqueness property for the sparse factorization, because the permutation ambiguity can be reduced to a *subgroup* \mathcal{P} of permutations. Let us formalize this idea in the following definition and proposition.

Definition 4.11 (PS-uniqueness of an EMF in Σ). *For any set $\Sigma \subseteq \mathbb{C}^{m \times r} \times \mathbb{C}^{n \times r}$ of pairs of factors and any subgroup \mathcal{P} of the group \mathcal{P}_r of permutation matrices of size $r \times r$, the pair $(\mathbf{X}, \mathbf{Y}) \in \Sigma$ is the PS-unique EMF of $\mathbf{Z} := \mathbf{X}\mathbf{Y}^\top$ in Σ , if, for any $(\mathbf{X}', \mathbf{Y}') \in \Sigma$ such that $\mathbf{X}'\mathbf{Y}'^\top = \mathbf{Z}$, there exists an invertible diagonal matrix $\mathbf{D} \in \mathcal{D}_r$ and a permutation matrix $\mathbf{P} \in \mathcal{P}$ such that $(\mathbf{X}', \mathbf{Y}') = (\mathbf{X}\mathbf{D}\mathbf{P}, \mathbf{Y}\mathbf{D}^{-1}\mathbf{P}^\top)$.*

Remark 4.12. When $\mathcal{P} = \mathcal{P}_r$, PS-uniqueness is equivalent to PS-uniqueness introduced in Definition 2.1. In particular, PS-uniqueness implies PS-uniqueness of an EMF.

Proposition 4.13. *Consider $N = 2^J$ and $2 \leq q \leq J$. Then, with $\mathbf{F}^{(\ell)}$ ($1 \leq \ell \leq J$) defined as in (3.6), the partial product $\mathbf{F}^{(q)} \dots \mathbf{F}^{(1)}$ admits a PS-unique EMF in the constraint set $\Sigma_{\Omega'_q}$, where Ω'_q is defined by (4.10), and $\mathcal{P} := \mathcal{P}_{2^{q-1}-\text{bd}} \subseteq \mathcal{P}_N$ is the subgroup of permutation matrices of size $N \times N$ which are block-diagonal with blocks of size $2^{q-1} \times 2^{q-1}$.*

Proof. Denote $\mathbf{X} := \mathbf{F}^{(q)}$, $\mathbf{Y} := (\mathbf{F}^{(q-1)} \dots \mathbf{F}^{(1)})^\top$, and $\mathbf{Z} := \mathbf{X}\mathbf{Y}^\top$. Let $(\mathbf{X}', \mathbf{Y}') \in \Sigma_{\Omega'_q}$ such that $\mathbf{X}'\mathbf{Y}'^\top = \mathbf{Z}$. Since (\mathbf{X}, \mathbf{Y}) is the PS-unique EMF of \mathbf{Z} in $\Sigma_{\Omega'_q}$, there exist $\mathbf{D} \in \mathcal{D}_r$ and $\mathbf{P} \in \mathcal{P}_N$ such that $(\mathbf{X}', \mathbf{Y}') = (\mathbf{X}\mathbf{D}\mathbf{P}, \mathbf{Y}\mathbf{D}^{-1}\mathbf{P}^\top)$. Our goal is to show $\mathbf{P} \in \mathcal{P}_{2^{q-1}-\text{bd}}$. By definition (4.10), $\text{supp}(\mathbf{Y}') \subseteq \mathbf{W}^{(q-1:1)}$, where $\mathbf{W}^{(q-1:1)}$ has been defined by (3.8). But by Lemma 4.7, $\text{supp}(\mathbf{Y}) = \mathbf{W}^{(q-1:1)}$. This means that:

$$\mathbf{W}^{(q-1:1)}\mathbf{P}^\top = \text{supp}(\mathbf{Y}\mathbf{D}^{-1}\mathbf{P}^\top) = \text{supp}(\mathbf{Y}') \subseteq \mathbf{W}^{(q-1:1)}.$$

In particular, this implies that \mathbf{P} is block-diagonal with blocks of size $2^{q-1} \times 2^{q-1}$. \blacksquare

This key proposition is sufficient to establish the hierarchical identifiability for the butterfly factorization of the DFT matrix, with relaxed sparsity constraint that we now precise. For any family of supports $\Lambda \subseteq \mathbb{B}^{m \times n}$, denote:

$$(4.11) \quad \Sigma_\Lambda := \bigcup_{\mathbf{S} \in \Lambda} \Sigma_{\mathbf{S}}, \quad \text{where } \Sigma_{\mathbf{S}} \text{ has been defined by (2.2).}$$

Recalling the notations (4.3)-(4.6), we consider the following sparsity constraint set:

$$(4.12) \quad \Sigma_{2-\text{col} \cap \text{bd}} := \Sigma_{\Lambda_J} \times \dots \times \Sigma_{\Lambda_1}, \quad \text{where } \Lambda_\ell := \Lambda_{2-\text{col}} \cap \Lambda_{2^\ell-\text{bd}}, \quad 1 \leq \ell \leq J.$$

Since this family introduces some permutation ambiguities, let us slightly extend [Definition 3.9](#) to take into account the permutation ambiguities in a multi-layer matrix factorization.

Definition 4.14 (PS-uniqueness of an MEMF in Σ). *The definition is analogue to the one of “S-uniqueness of an MEMF in Σ ” given by [Definition 3.9](#), where we replace invertible diagonal matrices by generalized permutation matrices.*

We are now able to formulate our final main contribution, which is the hierarchical identifiability (associated to an unbalanced partitioning binary tree) of the butterfly factorization of the DFT matrix with the relaxed constraint $\Sigma_{2\text{-co1}\cap\text{bd}}$.

Theorem 4.15. *Consider $N = 2^J$. Then, the matrix $\mathbf{Z} := \mathbf{F}^{(J)} \dots \mathbf{F}^{(1)}$, with $\mathbf{F}^{(\ell)}$ ($1 \leq \ell \leq J$) defined by [\(3.6\)](#), admits a PS-unique MEMF in $\Sigma_{2\text{-co1}\cap\text{bd}}$ defined by [\(4.12\)](#). Moreover, the factors $(\mathbf{F}^{(\ell)})_{\ell=1}^J$ can be recovered from \mathbf{Z} , up to scaling and permutation ambiguities only, through a hierarchical factorization method described by [Algorithm 4.1](#).*

Algorithm 4.1 Hierarchical factorization method for $\mathbf{DFT}_N \mathbf{R}_N^\top$, $N = 2^J$.

Require: Matrix $\mathbf{H}^{(J:1)} \in \mathbb{C}^{N \times N}$

- 1: **for** $q = J, J-1, \dots, 2$ **do**
 - 2: $(\mathbf{X}^{(q)}, \mathbf{H}^{(q-1:1)\top}) \leftarrow$ a solution to $\min_{(\mathbf{X}, \mathbf{Y}) \in \Sigma_{\Omega'_q}} \|\mathbf{H}^{(q:1)} - \mathbf{X}\mathbf{Y}^\top\|_F$
 - 3: **end for**
 - 4: Set $\mathbf{X}^{(1)} = \mathbf{H}^{(1:1)}$
 - 5: **return** $(\mathbf{X}^{(J)}, \mathbf{X}^{(J-1)}, \dots, \mathbf{X}^{(1)})$
-

Proof. Denote $\mathbf{Z} := \mathbf{F}^{(J)} \dots \mathbf{F}^{(1)}$. Let $(\mathbf{X}^{(J)}, \dots, \mathbf{X}^{(1)}) \in \Sigma_{2\text{-co1}\cap\text{bd}}$ such that $\mathbf{X}^{(J)} \dots \mathbf{X}^{(1)} = \mathbf{Z}$. For any $2 \leq q \leq J$, define the assertion P_q : “there exist $\mathbf{G}^{(J-1)}, \dots, \mathbf{G}^{(q-1)} \in \mathcal{G}_N$ with $\text{supp}(\mathbf{G}^{(\ell)}) \in \mathcal{P}_{2^\ell\text{-bd}}$ ($q-1 \leq \ell \leq J-1$) such that $\mathbf{X}^{(\ell)} = \mathbf{G}^{(\ell)-1} \mathbf{F}^{(\ell)} \mathbf{G}^{(\ell-1)}$ for any $\ell \in \{q, \dots, J\}$ and $\mathbf{X}^{(q-1)} \dots \mathbf{X}^{(1)} = \mathbf{G}^{(q-1)-1} \mathbf{F}^{(q-1)} \dots \mathbf{F}^{(1)}$ ”, where $\mathcal{P}_{2^\ell\text{-bd}}$ is the subgroup of permutation matrices which are block-diagonal with blocks of size $2^\ell \times 2^\ell$, and $\mathbf{G}^{(J)} := \mathbf{I}_N$ by convention. Showing P_q for $q = 2$ by backward induction would then yield our claim.

P_J can be verified by applying [Proposition 4.13](#) with $q = J$. Then, let $2 \leq q \leq J-1$, and suppose that P_{q+1} is true. Fix $\mathbf{G}^{(J-1)}, \dots, \mathbf{G}^{(q)} \in \mathcal{G}_N$ with $\text{supp}(\mathbf{G}^{(\ell)}) \in \mathcal{P}_{2^\ell\text{-bd}}$ ($q \leq \ell \leq J$) such that $\mathbf{X}^{(\ell)} = \mathbf{G}^{(\ell)-1} \mathbf{F}^{(\ell)} \mathbf{G}^{(\ell-1)}$ for any $\ell \in \{q+1, \dots, J\}$, and $\mathbf{X}^{(q)} \dots \mathbf{X}^{(1)} = \mathbf{G}^{(q)-1} \mathbf{F}^{(q)} \dots \mathbf{F}^{(1)}$, which is equivalent to $(\mathbf{G}^{(q)} \mathbf{X}^{(q)})(\mathbf{X}^{(q-1)} \dots \mathbf{X}^{(1)}) = \mathbf{F}^{(q)} \mathbf{F}^{(q-1)} \dots \mathbf{F}^{(1)}$. But $\mathbf{G}^{(q)}, \mathbf{X}^{(q)}$ are both block-diagonal with blocks of size $2^q \times 2^q$, and $\mathbf{X}^{(q-1)} \dots \mathbf{X}^{(1)}$ is block-diagonal with blocks of size $2^{q-1} \times 2^{q-1}$, hence: $(\mathbf{G}^{(q)} \mathbf{X}^{(q)}, (\mathbf{X}^{(q-1)} \dots \mathbf{X}^{(1)})^\top) \in \Sigma_{\Omega'_q}$. By [Proposition 4.13](#), there exists $\mathbf{G}^{(q-1)} \in \mathcal{G}_N$ with $\text{supp}(\mathbf{G}^{(q-1)}) \in \mathcal{P}_{2^{q-1}\text{-bd}}$ such that $\mathbf{G}^{(q)} \mathbf{X}^{(q)} = \mathbf{F}^{(q)} \mathbf{G}^{(q-1)}$ and $\mathbf{X}^{(q-1)} \dots \mathbf{X}^{(1)} = \mathbf{G}^{(q-1)-1} \mathbf{F}^{(q-1)} \dots \mathbf{F}^{(1)}$. Rearranging the former equality as $\mathbf{X}^{(q)} = \mathbf{G}^{(q)-1} \mathbf{F}^{(q)} \mathbf{G}^{(q-1)}$ yields P_q , which ends the proof by induction. \blacksquare

Remark 4.16. Although there does not exist yet a procedure that is guaranteed to solve the minimization problem at [Line 2](#) in [Algorithm 4.1](#), one possible heuristic is to apply the proximal gradient algorithm proposed by [\[12\]](#). Indeed, the proximal operator associated to the constraint set $\Sigma_{\Omega'_q}$ has a closed-form expression: it can be easily computed by setting

the entries outside the diagonal blocks to zero, and by keeping only the two largest entries in absolute value at each column.

5. Conclusion. We established hierarchical identifiability in the multi-layer sparse matrix factorization, when enforcing the butterfly structure as the sparsity constraint. We proved that the butterfly factors $\mathbf{X}^{(\ell)}$ of $\mathbf{Z} := \mathbf{X}^{(J)} \dots \mathbf{X}^{(1)}$ can be recovered from \mathbf{Z} by a hierarchical factorization method, up to scaling ambiguities. We also showed a stronger identifiability result for the butterfly factorization of the DFT matrix, by relaxing the sparsity constraint to only enforce 2-sparsity by column and a block-diagonal structure on each factor.

In complement to the research directions discussed at the end of the companion paper [20], we present some other perspectives in terms of algorithm and stability. Our analysis suggests other approaches for sparse matrix factorization than iterative first-order optimization methods [12, 3]. One can study the stability of Algorithm 3.1 with respect to noise and, as discussed in Remark 3.14, the empirical performance of Algorithm 3.1 to the noisy case, in which an exact factorization is not ensured to exist. In Algorithm 4.1, a future challenge is to find a robust and optimal procedure to solve the minimization problem at each hierarchical level in Line 2.

REFERENCES

- [1] S. CHOUDHARY AND U. MITRA, *Identifiability scaling laws in bilinear inverse problems*, arXiv preprint arXiv:1402.2637, (2014), <https://arxiv.org/abs/1402.2637>.
- [2] S. CHOUDHARY AND U. MITRA, *Sparse blind deconvolution: What cannot be done*, in 2014 IEEE International Symposium on Information Theory, IEEE, 2014, pp. 3002–3006, <https://doi.org/10.1109/ISIT.2014.6875385>.
- [3] T. DAO, A. GU, M. EICHHORN, A. RUDRA, AND C. RÉ, *Learning fast algorithms for linear transforms using butterfly factorizations*, in International Conference on Machine Learning, PMLR, 2019, pp. 1517–1527, <http://proceedings.mlr.press/v97/dao19a/dao19a.pdf>.
- [4] M. ELAD, *Sparse and redundant representations: from theory to applications in signal and image processing*, Springer Science & Business Media, 2010, <https://doi.org/10.1007/978-1-4419-7011-4>.
- [5] S. FOUCART AND H. RAUHUT, *A mathematical introduction to compressive sensing*, Applied and Numerical Harmonic Analysis, Birkhäuser Basel, 2013, <https://doi.org/10.1007/978-0-8176-4948-7>.
- [6] N. HALKO, P.-G. MARTINSSON, AND J. A. TROPP, *Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions*, SIAM review, 53 (2011), pp. 217–288, <https://doi.org/10.1137/090771806>.
- [7] Q.-T. LE, *Multilayer sparse matrix factorization*, tech. report, ENS Paris Saclay, Sept. 2020, <https://hal.inria.fr/hal-03130680>.
- [8] Q.-T. LE AND R. GRIBONVAL, *Structured support exploration for multilayer sparse matrix factorization*, in ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2021, pp. 3245–3249, <https://doi.org/10.1109/ICASSP39728.2021.9414238>.
- [9] Q.-T. LE, E. RICCIETTI, AND R. GRIBONVAL, *Spurious Valleys, Spurious Minima and NP-hardness of Sparse Matrix Factorization With Fixed Support*. preprint, <https://hal.inria.fr/hal-03364668>, May 2021, <https://hal.inria.fr/hal-03364668>.
- [10] L. LE MAGOAROU, *Matrices efficaces pour le traitement du signal et l'apprentissage automatique*, PhD thesis, INSA de Rennes, 2016, <https://hal.inria.fr/tel-01412558>. Written in French.
- [11] L. LE MAGOAROU AND R. GRIBONVAL, *Chasing butterflies: In search of efficient dictionaries*, in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2015, pp. 3287–3291, <https://doi.org/10.1109/ICASSP.2015.7178579>.
- [12] L. LE MAGOAROU AND R. GRIBONVAL, *Flexible multilayer sparse approximations of matrices and applications*, IEEE Journal of Selected Topics in Signal Processing, 10 (2016), pp. 688–700, <https://doi.org/10.1109/JSTSP.2016.2582444>.

- [//doi.org/10.1109/JSTSP.2016.2543461](https://doi.org/10.1109/JSTSP.2016.2543461).
- [13] Y. LI, K. LEE, AND Y. BRESLER, *Identifiability and stability in blind deconvolution under minimal assumptions*, IEEE Transactions on Information Theory, 63 (2017), pp. 4619–4633, <https://doi.org/10.1109/TIT.2017.2689779>.
- [14] F. MALGOUYRES, *On the stable recovery of deep structured linear networks under sparsity constraints*, in Mathematical and Scientific Machine Learning, PMLR, 2020, pp. 107–127, <http://proceedings.mlr.press/v107/malgouyres20a/malgouyres20a.pdf>.
- [15] F. MALGOUYRES AND J. LANDSBERG, *On the identifiability and stable recovery of deep/multi-layer structured matrix factorization*, in 2016 IEEE Information Theory Workshop (ITW), IEEE, 2016, pp. 315–319, <https://doi.org/10.1109/ITW.2016.7606847>.
- [16] F. MALGOUYRES AND J. LANDSBERG, *Multilinear compressive sensing and an application to convolutional linear networks*, SIAM Journal on Mathematics of Data Science, 1 (2019), pp. 446–475, <https://doi.org/10.1137/18M119834X>.
- [17] R. RUBINSTEIN, A. M. BRUCKSTEIN, AND M. ELAD, *Dictionaries for sparse representation modeling*, Proceedings of the IEEE, 98 (2010), pp. 1045–1057, <https://doi.org/10.1109/JPROC.2010.2040551>.
- [18] R. RUBINSTEIN, M. ZIBULEVSKY, AND M. ELAD, *Double sparsity: Learning sparse dictionaries for sparse signal approximation*, IEEE Transactions on Signal Processing, 58 (2010), pp. 1553–1564, <https://doi.org/10.1109/TSP.2009.2036477>.
- [19] C. F. VAN LOAN, *The ubiquitous kronecker product*, Journal of Computational and Applied Mathematics, 123 (2000), pp. 85–100, [https://doi.org/10.1016/S0377-0427\(00\)00393-9](https://doi.org/10.1016/S0377-0427(00)00393-9).
- [20] L. ZHENG, E. RICCIETTI, AND R. GRIBONVAL, *Identifiability in Two-Layer Sparse Matrix Factorization*, arXiv preprint arXiv:2110.01235, (2021), <https://arxiv.org/abs/2110.01235>.

Appendix A. Proof of Lemma 3.8.

Proof. The implication (i) \Rightarrow (ii) is direct by considering values of p, ℓ, q such that the partial products are made of a single factor. We now prove (ii) \Rightarrow (i). Suppose (ii). Fix $\ell \in \{2, \dots, J\}$, and let us show by induction that $\mathbf{H}^{(q:\ell)} := \mathbf{X}^{(q)} \dots \mathbf{X}^{(\ell)}$ has no zero column for each $q \in \{\ell, \dots, J\}$. By assumption, $\mathbf{H}^{(\ell:\ell)} = \mathbf{X}^{(\ell)}$ has no zero column. Let $q \in \{\ell, \dots, J-1\}$, and suppose that $\mathbf{H}^{(q:\ell)}$ has no zero column. Let $i \in \llbracket N \rrbracket$. Then, the i -th column of $\mathbf{H}^{(q+1:\ell)} = \mathbf{X}^{(q+1)} \mathbf{H}^{(q:\ell)}$ is a linear combination of columns in $\mathbf{X}^{(q+1)}$ indexed by $k \in \text{supp}((\mathbf{H}^{(q:\ell)})_i)$. By Lemma 3.7, the supports of the rank-one contributions in $\varphi(\mathbf{X}^{(q+1)}, \mathbf{H}^{(q:\ell)\top})$ are pairwise disjoint. In particular, the supports of the columns in $\mathbf{X}^{(q+1)}$ indexed by $k \in \text{supp}((\mathbf{H}^{(q:\ell)})_i)$ are pairwise disjoint. But $\text{supp}((\mathbf{H}^{(q:\ell)})_i)$ is not empty because by assumption $\mathbf{H}^{(q:\ell)}$ has no zero column, and $\mathbf{X}^{(q+1)}$ has also no zero column by assumption. Thus, the i -th column of $\mathbf{H}^{(q+1:\ell)}$ is not a zero column, and this is true for each $i \in \llbracket N \rrbracket$, which ends the recursion. A similar recursion shows that $\mathbf{X}^{(\ell)} \dots \mathbf{X}^{(p)}$ has no zero row for each $1 \leq p \leq \ell \leq J-1$. ■

Appendix B. Proof of Proposition 4.3.

Let us reuse some lemmas from [10, Chapter 7].

Lemma B.1 ([10, Lemma 4, Chapter 7]). *For any r -tuple \mathcal{C} of rank-one contributions, if $\sum_{i=1}^r \|\mathcal{C}^i\|_0 = \|\mathcal{A}(\mathcal{C})\|_0$, then the supports $\{\text{supp}(\mathcal{C}^i)\}_{i=1}^r$ are pairwise disjoint.*

Corollary B.2 ([10, Corollary 2, Chapter 7]). *Suppose there exists an integer M such that for all $\mathcal{S} \in \varphi(\Omega)$, $\|\mathcal{S}^i\|_0 \leq M$ for each $i \in \llbracket r \rrbracket$. Let $\mathcal{C} \in \Gamma_\Omega$ such that the cardinality of $\text{supp}(\mathcal{A}(\mathcal{C}))$ is rM . Then, the supports $\{\text{supp}(\mathcal{C}^i)\}_{i=1}^r$ are pairwise disjoint.*

Proof. We have $rM = \text{card}(\text{supp}(\mathcal{A}(\mathcal{C}))) \leq \sum_{i=1}^r \|\mathcal{C}^i\|_0 \leq \sum_{i=1}^r \|\mathcal{S}^i\|_0 \leq rM$, so all the terms of the sum are equal and we conclude by Lemma B.1. ■

Proof of Proposition 4.3. By construction, $\|\mathcal{S}^i\|_0 \leq \frac{N}{2} \times 2 = N$ for all $i \in \llbracket r \rrbracket$, $\mathcal{S} \in \varphi(\Omega)$,

and since $\|\mathbf{Z}\| = N^2$, the first condition of [Proposition 4.2](#) is verified, by direct application of [Corollary B.2](#). We now focus on the second condition. Let $\mathcal{S} \in \varphi^{\text{cp}}(\Omega)$ such that $\{\mathcal{S}^i\}_{i=1}^N$ is a partition of $\text{supp}(\mathbf{Z})$, and that the rank of $(\mathbf{Z}_{k,l})_{(k,l) \in \mathcal{S}^i}$ ($1 \leq i \leq r$) is at most one. Then, necessarily, the cardinality of each support \mathcal{S}^i ($1 \leq i \leq r$) is at least N , so \mathcal{S}^i has exactly $\frac{N}{2}$ nonzero rows and 2 nonzero columns. For each i , we write $\mathcal{S}^i := \{k_1^{(i)}, \dots, k_{N/2}^{(i)}\} \times \{l_1^{(i)}, l_2^{(i)}\}$ the indices of the rank-one support \mathcal{S}^i . The objective of the proof is to show that all the row indices $k_\alpha^{(i)}$ ($\alpha \in \llbracket \frac{N}{2} \rrbracket$) have the same parity, and that $l_1^{(i)} + l_2^{(i)} - 1 = N$.

B.1. Case of the DCT matrix. In this paragraph, we denote $\mathbf{Z} = \mathbf{DCT}_N$ defined by [\(4.1\)](#). For each i , the submatrix $(\mathbf{Z}_{k,l})_{(k,l) \in \mathcal{S}^i}$ is of rank one, so there exists a scalar $\lambda^{(i)}$ such that:

$$(B.1) \quad \cos \left[\frac{\pi}{N} \left(l_1^{(i)} - \frac{1}{2} \right) (k_\alpha^{(i)} - 1) \right] = \lambda^{(i)} \cos \left[\frac{\pi}{N} \left(l_2^{(i)} - \frac{1}{2} \right) (k_\alpha^{(i)} - 1) \right], \quad \forall \alpha \in \llbracket \frac{N}{2} \rrbracket.$$

Fix $i \in \llbracket N \rrbracket$. For simplicity, we omit the subscript (i) for the rest of the proof.

B.1.1. Case 1: odd row indices. Suppose that there exists α such that $k_\alpha = 1$. Let us show that k_α is odd for all $\alpha \in \llbracket \frac{N}{2} \rrbracket$, and $l_1 + l_2 - 1 = N$. Evaluating [\(B.1\)](#) at $k_\alpha = 1$ yields $\lambda = 1$. Thus, for each $\alpha \in \llbracket \frac{N}{2} \rrbracket$:

$$\begin{aligned} \frac{\pi}{N} \left(l_1 - \frac{1}{2} \right) (k_\alpha - 1) &\equiv \frac{\pi}{N} \left(l_2 - \frac{1}{2} \right) (k_\alpha - 1) \pmod{2\pi} \\ \text{or } \frac{\pi}{N} \left(l_1 - \frac{1}{2} \right) (k_\alpha - 1) &\equiv -\frac{\pi}{N} \left(l_2 - \frac{1}{2} \right) (k_\alpha - 1) \pmod{2\pi}, \end{aligned}$$

which is equivalent to:

$$(B.2) \quad (l_1 - l_2)(k_\alpha - 1) \equiv 0 \pmod{2N} \quad \text{or} \quad (l_1 + l_2 - 1)(k_\alpha - 1) \equiv 0 \pmod{2N}.$$

Suppose by contradiction that there exists α such that k_α is even. Then, $(k_\alpha - 1)$ is odd, so by [\(B.2\)](#) and Euclid's lemma, $2N$ divides either $(l_1 - l_2)$ or $(l_1 + l_2 - 1)$, because $2N$ is even. But both cases are not possible, because $(l_1 - l_2) \in \{-N + 1, \dots, -1\} \cup \{1, \dots, N - 1\}$, and $(l_1 + l_2 - 1) \in \{1, \dots, 2N - 1\}$. Therefore, all the indices k_α ($\alpha \in \llbracket \frac{N}{2} \rrbracket$) are odd. Suppose now by contradiction that $l_1 + l_2 - 1 \neq N$. Then, evaluating [\(B.2\)](#) for $k_\alpha = 3$, N divides either $(l_1 - l_2)$ or $(l_1 + l_2 - 1)$. Again, this is not possible, which shows that necessarily, $l_1 + l_2 - 1 = N$.

B.1.2. Case 2: even row indices. Suppose now there does not exist α such that $k_\alpha = 1$. Then, by the previous case, all the column indices k_α ($\alpha \in \llbracket \frac{N}{2} \rrbracket$) are odd, because the supports $\{\mathcal{S}^i\}_{i=1}^r$ are pairwise disjoint. Let us show that $l_1 + l_2 - 1 = N$. Evaluating [\(B.1\)](#) at $k_\alpha = 2$ yields:

$$\lambda = \frac{\cos \left[\frac{\pi}{N} \left(l_1 - \frac{1}{2} \right) \right]}{\cos \left[\frac{\pi}{N} \left(l_2 - \frac{1}{2} \right) \right]}.$$

Then, evaluating [\(B.1\)](#) at $k_\alpha = N$ yields:

$$\cos(aN - a) \cos(b) = \cos(a) \cos(bN - b)$$

where we denoted $a = \frac{\pi}{N} (l_1 - \frac{1}{2})$ and $b = \frac{\pi}{N} (l_2 - \frac{1}{2})$. This is equivalent to:

$$(B.3) \quad \cos(aN - a - b) + \cos(aN - a + b) = \cos(a - bN + b) + \cos(a + bN - b).$$

We compute:

$$(B.4) \quad \begin{aligned} \cos(aN - a - b) &= \begin{cases} \sin \left[\frac{\pi}{N} (l_1 + l_2 - 1) \right] & \text{if } l_1 \text{ is odd} \\ -\sin \left[\frac{\pi}{N} (l_1 + l_2 - 1) \right] & \text{otherwise} \end{cases}, \\ \cos(aN - a + b) &= \begin{cases} \sin \left[\frac{\pi}{N} (l_2 - l_1) \right] & \text{if } l_1 \text{ is odd} \\ -\sin \left[\frac{\pi}{N} (l_2 - l_1) \right] & \text{otherwise} \end{cases}, \\ \cos(a - bN + b) &= \begin{cases} \sin \left[\frac{\pi}{N} (l_1 + l_2 - 1) \right] & \text{if } l_2 \text{ is odd} \\ -\sin \left[\frac{\pi}{N} (l_1 + l_2 - 1) \right] & \text{otherwise} \end{cases}, \\ \cos(a + bN - b) &= \begin{cases} -\sin \left[\frac{\pi}{N} (l_2 - l_1) \right] & \text{if } l_2 \text{ is odd} \\ \sin \left[\frac{\pi}{N} (l_2 - l_1) \right] & \text{otherwise} \end{cases}. \end{aligned}$$

Suppose by contradiction that l_1 and l_2 have the same parity. Then (B.3) combined with (B.4) yields $\sin \left[\frac{\pi}{N} (l_2 - l_1) \right] = 0$, which means that N divides $l_1 - l_2$. This is impossible as $(l_1 - l_2) \in \{-N + 1, \dots, -1\} \cup \{1, \dots, N - 1\}$. Therefore, l_1 and l_2 do not have the same parity, and (B.3) combined with (B.4) yields $\sin \left[\frac{\pi}{N} (l_1 + l_2 - 1) \right] = 0$. Therefore, N divides $(l_1 + l_2 - 1)$, and since $(l_1 + l_2 - 1) \in \{1, \dots, 2N - 1\}$, we obtain $l_1 + l_2 - 1 = N$.

B.2. Case of the DST matrix. In this paragraph, we denote $\mathbf{Z} = \mathbf{DST}_N$ defined by (4.2). For each i , the submatrix $(\mathbf{Z}_{k,l})_{(k,l) \in \mathcal{S}^i}$ is of rank one, so there exists a scalar $\lambda^{(i)}$ such that:

$$(B.5) \quad \sin \left[\frac{\pi}{N} \left(l_1^{(i)} - \frac{1}{2} \right) k_\alpha^{(i)} \right] = \lambda^{(i)} \sin \left[\frac{\pi}{N} \left(l_2^{(i)} - \frac{1}{2} \right) k_\alpha^{(i)} \right], \quad \forall \alpha \in \left\lfloor \frac{N}{2} \right\rfloor.$$

Fix $i \in \llbracket N \rrbracket$. For simplicity, we omit the subscript (i) for the rest of the proof.

B.2.1. Case 1: even row indices. Suppose that there exists α such that $k_\alpha = N$. Let us show that k_α is even for all $\alpha \in \left\lfloor \frac{N}{2} \right\rfloor$, and $l_1 + l_2 - 1 = N$. Evaluating (B.5) at $k_\alpha = N$ yields:

$$\cos(\pi l_1) = \lambda \cos(\pi l_2).$$

When l_1 and l_2 have the same parity. In this case, $\lambda = 1$, and (B.5) yields for all $\alpha \in \left\lfloor \frac{N}{2} \right\rfloor$:

$$\begin{aligned} \frac{\pi}{N} \left(l_1 - \frac{1}{2} \right) k_\alpha &\equiv \frac{\pi}{N} \left(l_2 - \frac{1}{2} \right) k_\alpha \pmod{2\pi}, \\ \text{or } \frac{\pi}{N} \left(l_1 - \frac{1}{2} \right) k_\alpha &\equiv \pi - \frac{\pi}{N} \left(l_2 - \frac{1}{2} \right) k_\alpha \pmod{2\pi}, \end{aligned}$$

which simplifies into:

$$(B.6) \quad (l_1 - l_2)k_\alpha \equiv 0 \pmod{2N} \quad \text{or} \quad (l_1 + l_2 - 1)k_\alpha \equiv N \pmod{2N}, \quad \forall \alpha \in \left\lfloor \frac{N}{2} \right\rfloor.$$

Suppose by contradiction that there exists α such that k_α is odd. Since $(l_1 - l_2) \in \{-N + 1, \dots, -1\} \cup \{1, \dots, N - 1\}$, $2N$ does not divide $(l_1 - l_2)$. Consequently, $2N$ cannot divide $(l_1 - l_2)k_\alpha$, because $2N$ and k_α do not have the same parity. Therefore, by (B.6), $2N$ divides necessarily $(l_1 + l_2 - 1)k_\alpha - N$. But l_1 and l_2 have the same parity, so $(l_1 + l_2 - 1)$ is odd. As k_α is odd and N is even, the number $(l_1 + l_2 - 1)k_\alpha - N$ is odd, which is not possible. Therefore, we conclude that k_α is even for all $\alpha \in \llbracket \frac{N}{2} \rrbracket$. Suppose now by contradiction that $l_1 + l_2 - 1 \neq N$. By evaluating (B.6) at $k_\alpha = 2$, we obtain $(l_1 - l_2) \equiv 0 \pmod{N}$ or $(l_1 + l_2 - 1) \equiv \frac{N}{2} \pmod{N}$. The former is not possible, because $(l_1 - l_2) \in \{-N + 1, \dots, -1\} \cup \{1, \dots, N - 1\}$. The latter is also not possible, because $(l_1 + l_2 - 1)$ is odd, $\frac{N}{2}$ is even, and N is even. In conclusion, $l_1 + l_2 - 1 = N$.

When l_1 and l_2 do not have the same parity. In this case, $\lambda = -1$, and (B.5) yields for all $\alpha \in \llbracket \frac{N}{2} \rrbracket$:

$$\begin{aligned} \frac{\pi}{N} \left(l_1 - \frac{1}{2} \right) k_\alpha &\equiv -\frac{\pi}{N} \left(l_2 - \frac{1}{2} \right) k_\alpha \pmod{2\pi}, \\ \text{or } \frac{\pi}{N} \left(l_1 - \frac{1}{2} \right) k_\alpha &\equiv \pi + \frac{\pi}{N} \left(l_2 - \frac{1}{2} \right) k_\alpha \pmod{2\pi}, \end{aligned}$$

which simplifies into:

$$(B.7) \quad (l_1 + l_2 - 1)k_\alpha \equiv 0 \pmod{2N} \quad \text{or} \quad (l_1 - l_2)k_\alpha \equiv N \pmod{2N}, \quad \forall \alpha \in \llbracket \frac{N}{2} \rrbracket.$$

Suppose by contradiction that there exists α such that k_α is odd. Since $(l_1 + l_2 - 1) \in \{1, \dots, 2N - 1\}$, $2N$ does not divide $(l_1 + l_2 - 1)$. Hence, $2N$ cannot divide $(l_1 + l_2 - 1)k_\alpha$, because $2N$ and k_α do not have the same parity. Therefore, by (B.6), $2N$ divides necessarily $(l_1 - l_2)k_\alpha - N$. But l_1 and l_2 do not have the same parity, so $(l_1 - l_2)$ is odd. As k_α is odd and N is even, the number $(l_1 - l_2)k_\alpha - N$ is odd, which is not possible. Therefore, we conclude that k_α are even for all $\alpha \in \llbracket \frac{N}{2} \rrbracket$. Suppose now that by contradiction $l_1 + l_2 - 1 \neq N$. By evaluating (B.6) at $k_\alpha = 2$, we obtain $(l_1 + l_2 - 1) \equiv 0 \pmod{N}$ or $(l_1 - l_2) \equiv \frac{N}{2} \pmod{N}$. The former is not possible, because $(l_1 + l_2 - 1) \in \{1, \dots, N - 1\} \cup \{N + 1, \dots, 2N - 1\}$. The latter is also not possible, because $(l_1 - l_2)$ is odd, $\frac{N}{2}$ is even, and N is even. In conclusion, $l_1 + l_2 - 1 = N$.

B.2.2. Case 2: odd row indices. Suppose now that there does not exist α such that $k_\alpha = N$. Then, by the previous case, all the column indices k_α ($\alpha \in \llbracket \frac{N}{2} \rrbracket$) are even, because the supports $\{\mathcal{S}^{i'}\}_{i'=1}^r$ are pairwise disjoint. Let us show that $l_1 + l_2 - 1 = N$. Evaluating (B.5) at $k_\alpha = 1$ yields:

$$\lambda = \frac{\sin \left[\frac{\pi}{N} \left(l_1 - \frac{1}{2} \right) \right]}{\sin \left[\frac{\pi}{N} \left(l_2 - \frac{1}{2} \right) \right]}.$$

Then, evaluating (B.5) at $k_\alpha = N - 2$ yields:

$$\sin(aN - a) \sin(b) = \sin(a) \sin(bN - b)$$

where we denoted $a = \frac{\pi}{N} (l_1 - \frac{1}{2})$ and $b = \frac{\pi}{N} (l_2 - \frac{1}{2})$. This is equivalent to:

$$(B.8) \quad \cos(aN - a - b) - \cos(aN - a + b) = \cos(a - bN + b) - \cos(a + bN - b). \quad \blacksquare$$

Suppose by contradiction that l_1 and l_2 have the same parity. Then (B.8) combined with (B.4) yields $\sin[\frac{\pi}{N}(l_2 - l_1)] = 0$, which means that N divides $l_1 - l_2$. This is impossible as $(l_1 - l_2) \in \{-N + 1, \dots, -1\} \cup \{1, \dots, N - 1\}$. Therefore, l_1 and l_2 do not have the same parity, and (B.8) combined with (B.4) yields $\sin[\frac{\pi}{N}(l_1 + l_2 - 1)] = 0$. Therefore, N divides $(l_1 + l_2 - 1)$, and since $(l_1 + l_2 - 1) \in \{1, \dots, 2N - 1\}$, we obtain $l_1 + l_2 - 1 = N$.

Appendix C. Proof of Proposition 4.6.

Proof. By the definition of the Hadamard matrix (4.5), we have:

$$(\mathbf{H}_2 \otimes \mathbf{I}_{N/2})(\mathbf{I}_2 \otimes \mathbf{H}_{N/2}) \stackrel{(1.2)}{=} \mathbf{H}_2 \otimes \mathbf{H}_{N/2} \stackrel{(4.5)}{=} \mathbf{H}_N \stackrel{(4.5)}{=} \mathbf{H}_{N/2} \otimes \mathbf{H}_2 \stackrel{(1.2)}{=} (\mathbf{H}_{N/2} \otimes \mathbf{I}_2)(\mathbf{I}_{N/2} \otimes \mathbf{H}_2).$$

But the Hadamard matrix is symmetric, and $(\mathbf{A} \otimes \mathbf{B})^\top = \mathbf{A}^\top \otimes \mathbf{B}^\top$ for any \mathbf{A}, \mathbf{B} , so:

$$\mathbf{H}_N = \mathbf{H}_N^\top = (\mathbf{I}_2 \otimes \mathbf{H}_{N/2})^\top (\mathbf{H}_2 \otimes \mathbf{I}_{N/2})^\top = (\mathbf{I}_2 \otimes \mathbf{H}_{N/2})(\mathbf{H}_2 \otimes \mathbf{I}_{N/2})^\top.$$

In other words, the Hadamard matrix of size N admits the two following factorizations:

$$(\mathbf{I}_2 \otimes \mathbf{H}_{N/2})(\mathbf{H}_2 \otimes \mathbf{I}_{N/2})^\top = \mathbf{H}_N = (\mathbf{H}_{N/2} \otimes \mathbf{I}_2)(\mathbf{I}_{N/2} \otimes \mathbf{H}_2)^\top,$$

These two factorizations of \mathbf{H}_N satisfy the sparsity constraint Σ_Ω , because by definition of the Kronecker product, $\mathbf{I}_2 \otimes \mathbf{H}_{N/2}$, $\mathbf{H}_{N/2} \otimes \mathbf{I}_2$ are $N/2$ -sparse by column, and $\mathbf{H}_2 \otimes \mathbf{I}_{N/2}$, $\mathbf{I}_{N/2} \otimes \mathbf{H}_2$ are 2-sparse by column. However, by (3.8), $\text{supp}(\mathbf{I}_2 \otimes \mathbf{H}_{N/2}) = \mathbf{W}^{(J-1:1)}$, $\text{supp}(\mathbf{H}_{N/2} \otimes \mathbf{I}_2) = \mathbf{W}^{(J:2)}$, and $(\mathbf{W}^{(J-1:1)})_1 = \{1, \dots, N/2 - 1\}$ is different to $(\mathbf{W}^{(J:2)})_i$ for each $i \in \llbracket N \rrbracket$, because

$$(\mathbf{W}^{(J:2)})_i = \begin{cases} \{1, 3, 5, \dots, N - 1\} & \text{if } i \text{ is odd} \\ \{2, 4, 6, \dots, N\} & \text{otherwise} \end{cases}.$$

Hence, $(\mathbf{I}_2 \otimes \mathbf{H}_{N/2}, \mathbf{H}_2 \otimes \mathbf{I}_{N/2}) \not\sim (\mathbf{H}_{N/2} \otimes \mathbf{I}_2, \mathbf{I}_{N/2} \otimes \mathbf{H}_2)$, in the sense of permutation-scaling equivalence. This means that \mathbf{H}_N does not admit a PS-unique EMF in Σ_Ω . \blacksquare