



# **A lifting approach to ParaTuck-2 tensor decompositions**

Konstantin Usevich

## **► To cite this version:**

| Konstantin Usevich. A lifting approach to ParaTuck-2 tensor decompositions. 2024. <hal-03966869v3>

**HAL Id: hal-03966869**

**<https://hal.science/hal-03966869v3>**

Preprint submitted on 7 Mar 2025

**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.



HAL Authorization

# A LIFTING APPROACH TO PARATUCK-2 TENSOR DECOMPOSITIONS\*

KONSTANTIN USEVICH<sup>†</sup>

**Abstract.** The ParaTuck-2 decomposition (PT2D) of third-order tensor is a two-layer generalization of the well-known canonical polyadic decomposition (CPD). While being more flexible than the CPD, the PT2D also possesses similar uniqueness properties. In this paper, we show that under the best known uniqueness conditions, the exact PT2D can be computed by an algebraic algorithm (i.e., can the PT2D problems can be reduced to computing nullspaces and eigenvalues of certain matrices). We do so by lifting the slices of the tensor to higher-dimensional space, which also allows for refining the existing uniqueness conditions. The algorithms are developed for general PT2D and its symmetric version (DEDICOM), which leads to an algebraic algorithm for another generalization of the CPD, the PARAFAC2 decomposition. Our methods are also applicable in the approximation scenario, as shown by the numerical experiments.

**Key words.** tensor decomposition, ParaTuck-2, DEDICOM, PARAFAC2, uniqueness, lifting

**MSC codes.** 15A69, 15A23

**1. Introduction.** ParaTuck-2 decomposition [12] (or PT2D for short) decomposes a collection of  $K$  matrices  $\mathbf{T}_k \in \mathbb{F}^{I \times J}$ ,  $k = 1, \dots, K$ , (where  $\mathbb{F} = \mathbb{R}$  or  $\mathbb{C}$ ), jointly as a product

$$(1.1) \quad \mathbf{T}_k = \mathbf{A} \mathbf{D}_G^{(k)} \mathbf{F} \mathbf{D}_H^{(k)} \mathbf{B}^\top,$$

where the three matrix factors  $\mathbf{A} \in \mathbb{F}^{I \times R}$ ,  $\mathbf{B} \in \mathbb{F}^{J \times S}$ ,  $\mathbf{F} \in \mathbb{F}^{R \times S}$  are common for all  $\mathbf{T}_k$  and the middle factors  $\mathbf{D}_G^{(k)}$  and  $\mathbf{D}_H^{(k)}$ ,  $k = 1, \dots, K$  are  $R \times R$  and  $S \times S$  diagonal matrices, respectively, that depend on the index  $k$ . The PT2D is a decomposition of an  $I \times J \times K$  3rd-order tensor (whose frontal slices  $I \times J$  are the matrices  $\mathbf{T}_k$ ) and the decomposition problem is to recover the unknown factors given the  $\mathbf{T}_k$ .

The PT2D (1.1) can be viewed as a “2-layer” generalization of the usual joint matrix factorization ( $\mathbf{T}_k = \mathbf{A} \mathbf{D}^{(k)} \mathbf{B}^\top$ ,  $\mathbf{D}^{(k)}$  diagonal,  $k = 1, \dots, K$ ), thus the PT2D is a generalization of the well-known CP (canonical polyadic) decomposition [16]. The name ParaTuck-2 was suggested in [12] to show that PT2D shares the features of PARAFAC<sup>1</sup> and Tucker tensor decompositions [16].

PT2D appears in several applications. In multi-way analysis, the PT2D was proposed in the context psychometrics [12] and chemometrics [2] to account for additional variability in the joint factor analysis of several datasets. In the context of telecommunications, PT2D was introduced as an encoding scheme (as well as various generalizations [9]) in the context of multiple-input multiple-output wireless communications [3, 5, 17]. Recently, PT2D appeared in block-structured system identification and neural network approximation (in the frame of so-called decoupling approach [8]), where the CPD corresponds to models with 1 hidden layer [8] while the 2-layer case gives rise to the PT2D [4, 21]. The fully symmetric version of PT2D (case  $\mathbf{A} = \mathbf{B}$ ,  $\mathbf{D}_G^{(k)} = \mathbf{D}_H^{(k)}$  in (1.1)), also known as the three-way DEDICOM<sup>2</sup> [11], is a way to describe asymmetric relationships in multivariate data analysis (see [16]). DEDICOM is also tightly linked to the popular PARAFAC-2 model [18]; in particular, DEDICOM is used as a backbone in so-called indirect approaches to PARAFAC-2 decomposition [14].

One of the key features of the PT2D that makes it appealing in applications is its (essential) uniqueness properties (up to trivial permutations and scalings) (similarly to the CPD). As shown in [12], for the case  $R \leq I$ ,  $S \leq J$  (full column rank  $\mathbf{A}, \mathbf{B}$ ), and  $K$  is sufficiently large, the PT2D is essentially unique under mild (generic) conditions on the factors  $\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{D}_G^{(k)}, \mathbf{D}_H^{(k)}$ . A similar uniqueness result was also proved in [12] for a partially symmetric version of PT2D (a generalized DEDICOM), which also serves as a base for the best known uniqueness results of the PARAFAC-2 decomposition [19]. However, up to the author’s knowledge, no significant progress has been made on uniqueness of PT2D and DEDICOM/PARAFAC2 since [12].

\*Submitted to the editors DATE.

**Funding:** This work was supported by the ANR projects LeaFleT (ANR-19-CE23-0021-01) and AGDAM (ANR-23-CE94-0001).

<sup>†</sup>Université de Lorraine and CNRS, CRAN (Centre de Recherche en Automatique en Nancy), UMR 7039, Campus Sciences, BP 70239, 54506 Vandœuvre-lès-Nancy cedex, France ([konstantin.usevich@cnrs.fr](mailto:konstantin.usevich@cnrs.fr)).

<sup>1</sup>one of the older names for the CP decomposition

<sup>2</sup>The name is an abbreviation of “decomposition into directional components”

Despite the usefulness and the nice features of PT2D, reliable algorithms are lacking, especially in the non-symmetric case. As for local optimization algorithms, many standard optimization tools, such as ALS (alternating least squares [2, pp. 68–71]) suffer from slow convergence and local minima. In many applications, a simplifying assumption is introduced that at least one of the factors (e.g., example  $\mathbf{A}$ ) is known [5], [2, p. 213], which simplifies the problem. Several recent efforts focused on improving local optimization strategies, such as double contractions in ALS in [17] or stochastic optimization [22], but these methods still lack convergence proofs. PT2D can be also reformulated as a structured CPD [9], but this reformulation does not currently yield a reliable way to compute the decomposition (as the resulting tensor rank is too high). For DEDICOM (as well as for PARAFAC-2 [18]), the available algorithms mainly employ (local) alternating optimization minimization, such as ALS [13, 14] or ASALSAN<sup>3</sup> [1], and are based on efficient techniques for computing updates. Finally, the uniqueness results of [12] do not provide an explicit (algebraic) algorithm for PT2D or DEDICOM (unlike the algebraic algorithms to find the CPD under Kruskal’s uniqueness conditions [16, 7]).

In this paper, we show that, by lifting  $\mathbf{T}_k$  to a higher-dimensional space, we can design algebraic algorithms (relying on standard linear algebra tools) for the PT2D that work under the uniqueness conditions in [12] (i.e.  $R \leq I, S \leq J$  and  $K$  sufficiently large). We provide algebraic algorithms for the non-symmetric PT2D and DEDICOM (and as a consequence for PARAFAC2), which are first of this kind, up to the author’s knowledge. The lifting approach relies on properties of the kernel of a certain structured matrix constructed from  $\mathbf{T}_k$ . This approach also allows us to relax and clarify uniqueness results in [12]. The proofs are split into many intermediate steps and highlight the importance of each of the assumptions, as well as the structure of the core tensor.

The paper is structured as follows. Section 2 contains basic matrix and tensor notation used in the paper. Section 3 recalls known facts about PT2D, including ambiguities and a summary of the results of [12], and gives a preview of the results of the paper. Section 4 provides an alternative definition of the PT2D through the core tensor and triple product of matrices and discusses the properties of the core tensor that are crucial for the following. Sections 5 and 6 contain the main results and algorithms for the non-symmetric and the fully symmetric cases, respectively. In section 7, we discuss the advanced properties of the core tensor and its implications. Section 8 contains a preview of the proof for a special case  $R = S = 2$ . Sections 9 and 10 contain the proofs for the nonsymmetric and symmetric case, respectively. Finally, Section 11 contains details on implementation and the numerical experiments.

In the paper, for simplicity, we assume  $\mathbb{F} = \mathbb{R}$  (however, all the results can be easily translated to the case  $\mathbb{F} = \mathbb{C}$ ).

**2. Matrix and tensor notation.** Vectors, matrices, and tensors, will be respectively denoted by bold lowercase letters (e.g.  $\mathbf{u}$ ), with bold uppercase letters (e.g.  $\mathbf{M}$ ), and with bold calligraphic letters, (e.g.  $\mathcal{A}$ ). For simplicity, here vectors, matrices, and tensors, are viewed as 1-way, 2-way and d-way arrays respectively. Their corresponding entries will be denoted by  $\mathcal{A}_{ijk}$ ,  $M_{ij}$ , and  $u_i$ . The vectors  $\mathbf{v} \in \mathbb{F}^n$  in this paper are always thought of as column vectors (i.e. matrices  $\mathbb{F}^{n \times 1}$ ). The notation  $\mathbb{F}^{n_1 \times \dots \times n_d} = (\mathbb{F}^{n_1}) \otimes \dots \otimes (\mathbb{F}^{n_d})$  denotes the space of  $n_1 \times \dots \times n_d$  tensors (where  $\otimes$  stands for the tensor product).

For a tensor  $\mathcal{T} \in \mathbb{F}^{n_1 \times \dots \times n_d}$  we denote by  $\text{vec}\{\mathcal{T}\} \in \mathbb{F}^{n_1 \dots n_d}$  the standard column-major vectorization of the tensor. For example, for the matrix  $\mathbf{V} = [\mathbf{v}_1 \ \dots \ \mathbf{v}_J] \in \mathbb{F}^{I \times J}$ , the column-major vectorization stacks its columns

$$\text{vec}\{\mathbf{V}\} = [\mathbf{v}_1^\top \ \dots \ \mathbf{v}_J^\top]^\top$$

We also allow the vectorization to be applied to tensor subspaces, e.g.  $\text{vec}\{\mathcal{V}\} \subseteq \mathbb{F}^{IJK}$  for  $\mathcal{V} \subseteq \mathbb{F}^{I \times J \times K}$ . By  $\text{matr}_{I,J}\{\mathbf{v}\}$  we denote  $(I, J)$ -matricization of a vector  $\mathbf{v} \in \mathbb{F}^{I \times J}$  the matrix  $\mathbf{V}$  such that  $\mathbf{v} = \text{vec}\{\mathbf{V}\}$ . We will equally allow the matricization to be applied directly on tensors, that is  $\text{matr}_{m,n}\{\mathcal{T}\} = \text{matr}_{m,n}\{\text{vec}\{\mathcal{T}\}\}$ . We also denote by  $\text{tens}_{I,J,K}\{\mathbf{v}\} \in \mathbb{R}^{I \times J \times K}$  the tensorization of the vector  $\mathbb{R}^{IJK}$ .

We use  $\text{Span}\{\cdot\}$  to denote linear space spanned by a set of vectors (matrices, or tensors) and  $\text{range}\{\mathbf{Z}\}$  the range of a matrix. We use  $\text{lker}\{\mathbf{Z}\}$  to denote the left kernel (i.e., orthogonal complement of  $\text{range}\{\mathbf{Z}\}$ ). Operator  $\text{Diag}\{\cdot\}$  denotes the diagonal matrix constructed from a vector, or a block-diagonal matrix built from a sequence of matrices.

Operator  $\bullet_p$  denotes contraction on the  $p$ th index of a tensor; when contracted with a matrix, it is understood that summation is always performed on the second index of the matrix. For instance,

<sup>3</sup>alternating simultaneous approximation, least squares and Newton

$(\mathcal{A} \bullet_1 \mathbf{M})_{ijk} = \sum_\ell \mathcal{A}_{\ell jk} M_{i\ell}$ . We denote by  $\mathbf{T}^{(1)} \in \mathbb{F}^{I \times (JK)}$  and  $\mathbf{T}^{(2)} \in \mathbb{F}^{J \times (IK)}$ ,  $\mathbf{T}^{(3)} \in \mathbb{F}^{K \times (IJ)}$  the first, second and third unfoldings of a tensor  $\mathcal{T} \in \mathbb{F}^{I \times J \times K}$ , respectively. Note that  $\mathbf{T}^{(1)} = \text{matr}_{I,JK}\{\mathcal{T}\}$  and  $\mathbf{T}^{(3)} = (\text{matr}_{I,J,K}\{\mathcal{T}\})^\top$ .

The symbol  $\otimes$  is used for tensor (outer) product of vectors (or tensors). That is,  $\mathcal{T} = \mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}$  if  $\mathcal{T}_{ijk} = a_i b_j c_k$ . To avoid confusion, we use the symbol  $\boxtimes$  for the Kronecker product of matrices. The usual identity for vectorization of products of matrices holds,  $\text{vec}\{\mathbf{A}\mathbf{X}\mathbf{B}^\top\} = (\mathbf{B} \boxtimes \mathbf{A}) \text{vec}\{\mathbf{X}\}$ . This implies, that the vectorization reverses the order of products for rank-one vectors, we have  $\text{vec}\{\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{c}\} = \mathbf{c} \boxtimes \mathbf{b} \boxtimes \mathbf{a}$ . We use  $\odot$  for the Khatri-Rao product of matrices, i.e., for two matrices

$$(2.1) \quad \mathbf{A} = [\mathbf{a}_1 \quad \cdots \quad \mathbf{a}_R], \mathbf{B} = [\mathbf{b}_1 \quad \cdots \quad \mathbf{b}_R],$$

we have  $\mathbf{B} \odot \mathbf{A} = [\mathbf{b}_1 \boxtimes \mathbf{a}_1 \quad \cdots \quad \mathbf{b}_R \boxtimes \mathbf{a}_R]$ .

We will use the common notation

$$[\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \sum_{r=1}^R \mathbf{a}_r \otimes \mathbf{b}_r \otimes \mathbf{c}_r$$

to denote an  $R$ -term (canonical<sup>4</sup>) polyadic (CP) decomposition, or CPD, of an  $I \times J \times K$  tensor with  $\mathbf{A}, \mathbf{B}$  as in (2.1) and  $\mathbf{C} = [\mathbf{c}_1 \quad \cdots \quad \mathbf{c}_R]$ .

We view symmetric tensors as subspace of order- $d$   $n \times \cdots \times n$  tensors, denoted by  $S^d(\mathbb{R}^n) \subset \mathbb{R}^{n \times \cdots \times n}$ . For example  $S^2(\mathbb{R}^n)$  denotes the vector space of symmetric  $n \times n$  matrices. We denote by  $\sigma_d : \mathbb{R}^d \rightarrow \text{vec}\{S^d(\mathbb{R}^n)\}$  the operation that symmetrizes of the tensor given in vectorized form (if is also an orthogonal projection on  $\text{vec}\{S^d(\mathbb{R}^n)\}$ ). For example, the symmetrization map  $\sigma_2 : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$  maps  $\text{vec}\{\mathbf{X}\} \mapsto \text{vec}\{\frac{\mathbf{X} + \mathbf{X}^\top}{2}\}$ . In particular,  $\sigma_2(\mathbf{y} \boxtimes \mathbf{z}) = \frac{1}{2}(\mathbf{y} \boxtimes \mathbf{z} + \mathbf{z} \boxtimes \mathbf{y})$  for  $\mathbf{y}, \mathbf{z} \in \mathbb{R}^n$ . Finally, for a linear subspace  $\mathcal{V} \subset \mathbb{R}^n$  we denote by  $S^2(\mathcal{V}) \subset \mathbb{R}^{n \times n}$  the set of symmetric matrices with range in  $\mathcal{V}$ .

### 3. Background on PT2D and its variants.

**3.1. ParaTuck-2 decomposition: notation and basic considerations.** We first introduce the following compact notation for the factors of the PT2D (that groups the diagonal factors  $\mathbf{D}_G^{(k)}$  and  $\mathbf{D}_H^{(k)}$  into two dense matrices  $\mathbf{F}$  and  $\mathbf{G}$ ).

**DEFINITION 3.1.** A third-order tensor  $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$  is said to admit a rank- $(R, S)$  Paratuck-2 decomposition with factors

$$\mathbf{A} \in \mathbb{F}^{I \times R}, \mathbf{B} \in \mathbb{R}^{J \times S}, \mathbf{F} \in \mathbb{R}^{R \times S}, \mathbf{G} \in \mathbb{R}^{R \times K}, \mathbf{H} \in \mathbb{R}^{S \times K},$$

if its frontal slices can be expressed as  $\mathcal{T}_{:,:,k} = \mathbf{T}_k$  with  $\mathbf{T}_k$  given in (1.1), where the  $\mathbf{D}_G^{(k)}$  and  $\mathbf{D}_H^{(k)}$  are defined from the columns of

$$(3.1) \quad \mathbf{G} = [\mathbf{g}_1 \quad \cdots \quad \mathbf{g}_K], \quad \mathbf{H} = [\mathbf{h}_1 \quad \cdots \quad \mathbf{h}_K]$$

as follows:

$$(3.2) \quad \mathbf{D}_G^{(k)} = \text{Diag}\{\mathbf{g}_k\}, \quad \mathbf{D}_H^{(k)} = \text{Diag}\{\mathbf{h}_k\}.$$

We will write use the notation

$$\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}),$$

to denote that  $\mathcal{T}$  has a PT2D with factors  $\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}$ .

**Example 3.2.** As mentioned in the introduction, CPD is a special case of PT2D for a particular choice of factors. Indeed, for  $R = S$ ,  $\mathbf{1}_{R \times K}$  denoting an  $R \times K$  matrix of all ones, we have

$$[\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!] = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{I}, \mathbf{C}, \mathbf{1}_{R \times K}).$$

Note that there is a plenty of ways to express a CPD as a special case of the PT2D.

**Remark 3.3.** Unlike the tensor rank (minimal  $R$  such that the tensor admits an  $R$ -term CPD), for PT2D, we cannot define the “minimal” rank in a unique way, because there is a pair  $(R, S)$  of “ranks”. In what follows, we will use the term “rank- $(R, S)$ ” ParaTuck-2 decomposition (or  $(R, S)$ -PT2D) without assuming that this is a minimal rank.

<sup>4</sup>We will not make a distinction between polyadic and canonical polyadic decomposition (in some articles, the term “canonical” is used only for  $[\![\mathbf{A}, \mathbf{B}, \mathbf{C}]\!]$  where  $R$  is the minimal possible such  $R$ ).

**3.2. Symmetric versions of PT2D and DEDICOM.** Several special cases of PT2D exist (see the introduction in [12]) were proposed:

- symmetrically weighted PT2D [12] :  $R = S$ ,  $\mathbf{G} = \mathbf{H}$  (which corresponds to  $\mathbf{D}_G^{(k)} = \mathbf{D}_H^{(k)}$  in (1.1));
- fully symmetric case (also known as DEDICOM [12, 16] ):  $R = S$ ,  $\mathbf{G} = \mathbf{H}$ ,  $\mathbf{A} = \mathbf{B}$  and  $\mathbf{F}$  symmetric. We will use the notation

$$\text{DEDICOM}(\mathbf{A}, \mathbf{F}, \mathbf{G}) := \text{PT2D}(\mathbf{A}, \mathbf{A}, \mathbf{F}, \mathbf{G}, \mathbf{G}).$$

*Remark 3.4.* For any matrix  $\mathbf{F}$ , we have  $\text{DEDICOM}(\mathbf{A}, \mathbf{F}, \mathbf{G}) = \text{DEDICOM}(\mathbf{A}, (\mathbf{F} + \mathbf{F}^\top)/2, \mathbf{G})$ . Therefore it makes no sense to consider nonsymmetric  $\mathbf{F}$  in DEDICOM.

The fully symmetric is it tightly linked to the PARAFAC-2 decomposition, as explained below.

*Example 3.5.* Let  $\mathbf{X}_k \in \mathbb{R}^{I \times J_k}$ ,  $k = 1, \dots, K$  be a collection of matrices that are decomposed as

$$(3.3) \quad \mathbf{X}_k = \mathbf{A} \mathbf{D}_k \mathbf{B}_k^\top,$$

where  $\mathbf{A} \in \mathbb{R}^{I \times R}$ ,  $\mathbf{D}_k \in \mathbb{R}^{R \times R}$  and  $\mathbf{B}_k \in \mathbb{R}^{R \times J_k}$  satisfy  $\mathbf{B}_k \mathbf{B}_k^\top = \mathbf{\Sigma}$ . Such a decomposition is known as PARAFAC-2 decomposition (with common covariance constraint).

PARAFAC-2 decomposition is linked to DEDICOM as follows. Under the common covariance constraint, the matrices  $\mathbf{M}_k = \mathbf{X}_k \mathbf{X}_k^\top$  can be expressed as

$$\mathbf{M}_k = \mathbf{A} \mathbf{D}_k \mathbf{\Sigma} \mathbf{D}_k \mathbf{A}^\top$$

Thus the “covariance” tensor  $\mathcal{M} \in \mathbb{R}^{I \times I \times K}$  with  $\mathcal{M}_{:, :, k} = \mathbf{M}_k$  admits the rank- $R$  fully symmetric PT2D ( $\mathcal{M} = \text{DEDICOM}(\mathbf{A}, \mathbf{\Sigma}, \mathbf{G})$ ).

The relation between DEDICOM and PARAFAC-2 decomposition in Example 3.5 is in fact used to establish the uniqueness properties of PARAFAC-2 [19], based on uniqueness proofs for DEDICOM in [12]. Also, DEDICOM is used to compute the PARAFAC-2 decomposition using the indirect approach [14].

**3.3. Ambiguities and uniqueness.** In this section, we will recall the trivial ambiguities of PT2D (which are similar to the ones of the CPD) and define the notion of (essential) uniqueness.

As in the case of the CPD, there are permutation and scaling ambiguities. For example, permuting the columns of  $\mathbf{A}$  together with the same permutation of rows  $\mathbf{F}$  and  $\mathbf{G}$ , leads to an alternative PT2D. We will summarize these ambiguities in the following lemma.

**LEMMA 3.6.** *The following transformations (trivial ambiguities) give an alternative PT2D:*

1.  $\text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \text{PT2D}(\mathbf{A} \mathbf{\Pi}_A, \mathbf{B}, \mathbf{\Pi}_A^\top \mathbf{F}, \mathbf{\Pi}_A^\top \mathbf{G}, \mathbf{H})$ , for any permutation matrix  $\mathbf{\Pi}_A \in \mathbb{R}^{R \times R}$
2.  $\text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \text{PT2D}(\mathbf{A}, \mathbf{B} \mathbf{\Pi}_B, \mathbf{F} \mathbf{\Pi}_B, \mathbf{G}, \mathbf{\Pi}_B^\top \mathbf{H})$ , for any permutation matrix  $\mathbf{\Pi}_B \in \mathbb{R}^{S \times S}$ .
3.  $\text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \text{PT2D}(\mathbf{A} \mathbf{\Lambda}_A, \mathbf{B}, (\mathbf{\Lambda}_A \mathbf{\Lambda}_G)^{-1} \mathbf{F}, \mathbf{\Lambda}_G \mathbf{G}, \mathbf{H})$ , for all invertible diagonal  $\mathbf{\Lambda}_A, \mathbf{\Lambda}_G$ .
4.  $\text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \text{PT2D}(\mathbf{A}, \mathbf{B} \mathbf{\Lambda}_B, \mathbf{F} (\mathbf{\Lambda}_B \mathbf{\Lambda}_H)^{-1}, \mathbf{G}, \mathbf{\Lambda}_H \mathbf{H})$ , for all invertible diagonal  $\mathbf{\Lambda}_B, \mathbf{\Lambda}_H$ .
5.  $\text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G} \mathbf{\Lambda}_{GH}, \mathbf{H} \mathbf{\Lambda}_{GH}^{-1})$ , for all invertible diagonal  $\mathbf{\Lambda}_{GH}$ .

*Proof.* Follows from explicit computations.  $\square$

We will call the PT2D unique if there are no other other PT2Ds rather than those obtained from Lemma 3.6.

**DEFINITION 3.7.** *The decomposition  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$  is called (essentially) unique if for any alternative decomposition*

$$\mathcal{T} = \text{PT2D}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \tilde{\mathbf{H}}),$$

*can be obtained through a sequence of transformations described in Lemma 3.6, i.e., there exist permutation matrices  $\mathbf{\Pi}_A \in \mathbb{R}^{R \times R}$  and  $\mathbf{\Pi}_B \in \mathbb{R}^{S \times S}$  and nonsingular diagonal matrices*

$$\mathbf{\Lambda}_A, \mathbf{\Lambda}_G \in \mathbb{R}^{R \times R}, \mathbf{\Lambda}_B, \mathbf{\Lambda}_H \in \mathbb{R}^{S \times S}, \mathbf{\Lambda}_{GH} \in \mathbb{R}^{K \times K}$$

such that

$$\begin{aligned}\tilde{\mathbf{A}} &= \mathbf{A} \cdot (\Pi_{\mathbf{A}} \cdot \Lambda_{\mathbf{A}}), & \tilde{\mathbf{B}} &= \mathbf{B} \cdot (\Pi_{\mathbf{B}} \cdot \Lambda_{\mathbf{B}}), \\ \tilde{\mathbf{F}} &= \Pi_{\mathbf{A}}^T \cdot (\Lambda_{\mathbf{G}}^{-1} \cdot \mathbf{F} \cdot \Lambda_{\mathbf{H}}^{-1}) \cdot \Pi_{\mathbf{B}}, \\ \tilde{\mathbf{G}} &= (\Lambda_{\mathbf{A}}^{-1} \cdot \Pi_{\mathbf{A}}^T) \cdot (\Lambda_{\mathbf{G}} \cdot \mathbf{G} \cdot \Lambda_{\mathbf{GH}}), \\ \tilde{\mathbf{H}} &= (\Lambda_{\mathbf{B}}^{-1} \cdot \Pi_{\mathbf{B}}^T) \cdot (\Lambda_{\mathbf{H}} \cdot \mathbf{H} \cdot \Lambda_{\mathbf{GH}}^{-1}).\end{aligned}$$

In the next subsection, we recall the existing results on uniqueness of PT2D.

**3.4. Known results: nonsymmetric case.** The uniqueness results of [12] rely on the following assumptions.

ASSUMPTION 1. The matrices  $\mathbf{A}$  and  $\mathbf{B}$  are of full column rank (i.e.  $\text{rank}\{\mathbf{A}\} = R, \text{rank}\{\mathbf{B}\} = S$ ).

ASSUMPTION 2. The matrix  $\mathbf{F}$  does not have zero elements.

ASSUMPTION 3. The matrix  $\mathbf{F}$  is maximal possible rank (i.e.  $\text{rank}\{\mathbf{F}\} = \min(R, S)$ ).

ASSUMPTION 4. All elements of  $\mathbf{G}$  and  $\mathbf{H}$  are non-zero (equivalently, all  $\mathbf{D}_{\mathbf{G}}^{(k)}, \mathbf{D}_{\mathbf{H}}^{(k)}$  are nonsingular).

ASSUMPTION 5. The matrix  $\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}$  has the maximal possible rank

$$\text{rank}\{\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}\} = \frac{R(R+1)}{2} \frac{S(S+1)}{2}.$$

Then the following theorem holds true

**THEOREM 3.8** ([12, Theorem 1]). Let  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$  be an  $(R, S)$ -PT2D whose factors satisfy [Assumptions 1 to 5](#). Then the PT2D is unique in the sense of [Definition 3.7](#).

The assumptions are quite natural (except probably [Assumption 5](#)); we will explain the intuition behind [Assumption 5](#) in the next sections. We will also show that, using our approach, [Assumptions 3](#) and [4](#) can be relaxed. Moreover, all these assumptions are generic as long as the matrices have certain dimensions.

*Remark 3.9* (On genericity of assumptions). [Assumptions 2](#) to [4](#) are satisfied for generic  $\mathbf{F}, \mathbf{G}, \mathbf{H}$ . i.e., the assumptions are satisfied except a set of  $\mathbf{F}, \mathbf{G}, \mathbf{H}$  of Lebesgue measure 0<sup>5</sup> (equivalently all the assumptions are satisfied with probability 1 for factors drawn from any absolutely continuous probability distribution.). If  $I \geq R, J \geq S$ , then [Assumption 1](#) is also satisfied for generic  $\mathbf{A}$  and  $\mathbf{B}$ , while [Assumption 5](#) is satisfied generically if  $K \geq \frac{R(R+1)}{2} \frac{S(S+1)}{2}$ .

*Remark 3.10.* Note that  $I \geq R, J \geq S$  is a necessary condition for [Assumption 1](#) while  $K \geq \frac{R(R+1)}{2} \frac{S(S+1)}{2}$  is necessary for [Assumption 5](#). ■

**3.5. Known results: symmetrically weighted case.** The symmetrically weighted case is slightly different from the non-symmetric case: [Theorem 3.8](#) does not apply, because [Assumption 5](#) is never satisfied. Indeed, if  $\mathbf{G} = \mathbf{H}$ , then  $\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H} = \mathbf{G}^{\odot 4}$ , whose rank is at most  $\binom{R+3}{4}$  which is strictly less than  $\binom{R+1}{2}^2$  (the bound in [Assumption 5](#)). Therefore, [Assumption 5](#) is replaced by the following assumption.

ASSUMPTION 6. It holds that  $\mathbf{G} = \mathbf{H}$  and the matrix  $\mathbf{G} \odot \mathbf{G} \odot \mathbf{G} \odot \mathbf{G}$  has the maximal possible rank

$$\text{rank}\{\mathbf{G} \odot \mathbf{G} \odot \mathbf{G} \odot \mathbf{G}\} = \binom{R+3}{4} = \frac{(R+3)(R+2)(R+1)R}{24}.$$

Then the following result holds true.

**THEOREM 3.11** ([12, Theorem 2]). Let  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{G})$  be the symmetrically weighted PT2D (with  $R = S$ ) whose factors satisfy [Assumptions 1](#) to [4](#) and [Assumption 6](#). Then the symmetrically weighted PT2D is unique in the sense of [Definition 3.7](#), i.e. any alternative symmetrically weighted PT2D

$$\mathcal{T} = \text{PT2D}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \tilde{\mathbf{G}}),$$

is necessarily as in [Definition 3.7](#) (with an extra constraint  $\tilde{\mathbf{G}} = \tilde{\mathbf{H}}$  and  $\Pi_{\mathbf{A}} = \Pi_{\mathbf{B}}$ ).

<sup>5</sup>More precisely, the exceptional set is a semialgebraic subset of strictly smaller dimension



*Remark 3.12* (On genericity of the assumptions). The [Assumption 6](#) is again satisfied for generic  $\mathbf{G}$  if  $K \geq \frac{R(R+1)(R+2)(R+3)}{24}$ . Note that  $K \geq \frac{R(R+1)(R+2)(R+3)}{24}$  is a necessary condition for [Assumption 6](#) to hold.

In particular, [Theorem 3.11](#) also applies to the fully symmetric PT2D (DEDICOM), and is used to establish uniqueness of PARAFAC-2 [15] (thanks to the correspondence in [Example 3.5](#)). Up to the author's knowledge, [Theorem 3.11](#) yields the best known results for uniqueness of DEDICOM (and, as a consequence, of the PARAFAC-2 decomposition), except a very particular case in [19], where generic uniqueness is shown for  $R = 2$ ,  $K = 4$  (see also references in [18]).

**3.6. Preview of the results of the paper.** The drawback of the results of [12] is that they do not come with an algorithm that can recover the factors of a given PT2D under the assumptions of [Theorem 3.8](#) and [Theorem 3.11](#). In this paper, we fill this gap by providing the algorithms for the non-symmetric PT2D and DEDICOM (the case of symmetrically weighted PT2D is omitted, but can be also treated along the same lines). Moreover, we show that some assumptions can be relaxed.

First of all [Assumption 4](#) can be completely removed (if [Assumption 5](#) or [Assumption 6](#) are satisfied). Second, [Assumption 3](#) can be relaxed to the following assumption.

ASSUMPTION 7.  $\text{krank}\{\mathbf{F}^\top\} \geq 2$  and  $\text{krank}\{\mathbf{F}\} \geq 2$ .

[Assumption 7](#) actually means that  $\mathbf{F}$  has neither proportional columns nor proportional rows. Then, in particular, the following two results hold true.

PROPOSITION 3.13 ([Proposition 5.5](#)). Let  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) \in \mathbb{R}^{I \times J \times K}$ , whose factors satisfy [Assumptions 1, 2, 5 and 7](#). Then the  $(R, S)$ -PT2D of  $\mathcal{T}$  is essentially unique. Also,  $(R, S)$  are minimal ranks (i.e., there does not exist  $(R', S')$ -PT2D with either  $R' < R$  or  $S' < S$ ).

PROPOSITION 3.14 ([Proposition 6.5](#)). Let  $\mathcal{T} = \text{DEDICOM}(\mathbf{A}, \mathbf{F}, \mathbf{G}) \in \mathbb{R}^{I \times J \times K}$ ,  $\mathbf{F} \in \mathbb{R}^{R \times R}$  symmetric, whose factors satisfy [Assumptions 1, 2, 6 and 7](#). Then the  $R$ -DEDICOM of  $\mathcal{T}$  is essentially unique. In addition,  $R$  is the minimal possible rank for a DEDICOM of  $\mathcal{T}$ .

Our approach relies on constructing the so-called “lifting” matrix  $\Phi(\mathcal{T})$ , whose left kernel has a particular structure. This enables us to retrieve the factors  $\mathbf{A}$  and  $\mathbf{B}$  from the left kernel of the lifting matrix (up to permutation and scaling of columns), uniquely, under the assumptions mentioned above. The other factors ( $\mathbf{F}$ ) can be retrieved uniquely (up to remaining ambiguities) once  $\mathbf{A}$  and  $\mathbf{B}$  are fixed.

The lifting approach provides an algorithm to recover the factors  $(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$  for the PT2D and  $\mathbf{A}, \mathbf{F}, \mathbf{G}$  for DEDICOM), up to trivial ambiguities, and also implies uniqueness of the decompositions. We also show that if [Assumption 7](#) is not satisfied, this leads to additional ambiguities of the PT2D, namely  $\mathbf{A}$  (resp.  $\mathbf{B}$ ) is non-unique if  $\text{krank}\{\mathbf{F}^\top\} = 2$  (resp.  $\text{krank}\{\mathbf{F}\} \geq 2$ ). However, in that case, the set of equivalent solutions can be described.

Although we do not cover the symmetrically weighted case in this paper, it can be treated along the same lines by applying the lifting approach.

**4. PT2D and its core tensor.** In this section, we recall an alternative representation for the PT2D via the core tensor (see [9]) and discuss the basic implications.

**4.1. Triple product of matrices.** First, we define the following special product of three matrices.

DEFINITION 4.1 (Triple product). For three matrices  $\mathbf{F} \in \mathbb{R}^{R \times S}$ ,  $\mathbf{G} \in \mathbb{R}^{R \times K}$ ,  $\mathbf{H} \in \mathbb{R}^{S \times K}$ , their triple product, denoted as  $\mathcal{C} = \text{TP}(\mathbf{F}, \mathbf{G}, \mathbf{H})$ , is the tensor  $\mathcal{C} \in \mathbb{R}^{R \times S \times K}$  whose elements are defined by

$$(4.1) \quad \mathcal{C}_{ijk} = F_{ij} G_{ik} H_{jk}.$$

Graphically, [Definition 4.1](#) can be visualised as a “product” of three matrices, see [Figure 4.1](#).

Fig. 4.1: Product of three matrices.

*Remark 4.2.* If  $\mathcal{C} = \text{TP}(\mathbf{F}, \mathbf{G}, \mathbf{H})$ , then its slices are equal to

$$(4.2) \quad \mathcal{C}_{::,k} = (\mathbf{F} \otimes \mathbf{g}_k \mathbf{h}_k^\top),$$

where  $\mathbf{g}_k, \mathbf{h}_k$  are as in (3.1), and  $\otimes$  is the Hadamard product (i.e., slices of  $\mathcal{C}$  are Hadamard products of the same matrix  $\mathbf{F}$  with different rank-one matrices).

*Remark 4.3.* The tensor factorization (4.1) is known in algebraic statistics as the “no-three-way interaction” model [6]. It is a typical example of a tensor factorization that cannot be realized as a graphical model.

**4.2. Factorizing the core tensor.** We also remark that the factorization  $\mathcal{C} = \text{TP}(\mathbf{F}, \mathbf{G}, \mathbf{H})$  admits trivial ambiguities

LEMMA 4.4. *A tensor  $\mathcal{C} = \text{TP}(\mathbf{F}, \mathbf{G}, \mathbf{H})$  admits an alternative decomposition  $\mathcal{C} = \text{TP}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \tilde{\mathbf{H}})$  for all  $\tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \tilde{\mathbf{H}}$  as*

$$\tilde{\mathbf{F}} = \Lambda_{\mathbf{G}}^{-1} \mathbf{F} \Lambda_{\mathbf{H}}^{-1}, \quad \tilde{\mathbf{G}} = \Lambda_{\mathbf{G}} \mathbf{G} \Lambda_{\mathbf{GH}}, \quad \tilde{\mathbf{H}} = \Lambda_{\mathbf{H}} \mathbf{H} \Lambda_{\mathbf{GH}}^{-1},$$

where  $\Lambda_{\mathbf{G}} \in \mathbb{R}^{R \times R}$ ,  $\Lambda_{\mathbf{H}} \in \mathbb{R}^{S \times S}$ ,  $\Lambda_{\mathbf{GH}} \in \mathbb{R}^{K \times K}$  are any diagonal nonsingular matrices.

*Proof.* Follows from straightforward calculation.  $\square$

The basic question is, whether for a tensor  $\mathcal{C}$  as in (4.1), we can recover the factors  $\mathbf{F}, \mathbf{G}, \mathbf{H}$ , given  $\mathcal{C}$ . We show that  $\mathcal{C}$  can be uniquely factorized (up to the ambiguities in Lemma 4.4) under some additional conditions. We propose the following two conditions that lead to two different ways to compute the factors.

LEMMA 4.5. *Let  $\mathcal{C}$  be of the form (4.1). If there exists one index  $r$  such that  $\mathcal{C}_{ijr} \neq 0$  for any  $i, j$ , then  $\mathbf{F}, \mathbf{G}, \mathbf{H}$  can be uniquely recovered from  $\mathcal{C}$  (up to scaling ambiguities in Lemma 4.4).*

*Remark 4.6.* In terms of assumptions on factors, the condition of Lemma 4.5 is equivalent to requiring that the Assumption 2 ( $F_{ij} \neq 0$ ) holds together with  $\mathbf{g}_r$  and  $\mathbf{h}_r$  not having zero elements for a fixed  $r$  (i.e.,  $\mathbf{D}_{\mathbf{G}}^{(r)}$  and  $\mathbf{D}_{\mathbf{H}}^{(r)}$  are nonsingular for fixed  $r$ ).

However, the factors of the core tensors can be recovered even if none of  $\mathbf{D}_{\mathbf{G}}^{(r)}$  and  $\mathbf{D}_{\mathbf{H}}^{(r)}$  are simultaneously nonsingular, but under an additional assumption.

PROPOSITION 4.7. *Let  $\mathcal{C}$  be of the form (4.1), whose factors  $\mathbf{F}, \mathbf{G}, \mathbf{H}$  satisfy Assumption 2 and either of Assumptions 5 to 6. Then  $\mathbf{F}, \mathbf{G}, \mathbf{H}$  can be uniquely recovered (up to the scaling ambiguities in Lemma 4.4) from  $\mathcal{C}$ .*

The proofs of Lemma 4.5 and Proposition 4.7 are constructive (i.e., provide algorithms to compute the factorization) and are given in Subsection 7.2.

**4.3. PT2D through the triple product.** It is easy to show that the PT2D can be defined via the triple product of matrices.

LEMMA 4.8. *For any  $\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}$  as in Definition 3.1 we have*

$$(4.3) \quad \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) = \left( \text{TP}(\mathbf{F}, \mathbf{G}, \mathbf{H}) \right) \bullet_1 \mathbf{A} \bullet_2 \mathbf{B}.$$

*Proof.* Indeed if  $\mathcal{C} = \text{TP}(\mathbf{F}, \mathbf{G}, \mathbf{H})$ , then it is easy to see from (4.2) that its slices satisfy

$$(4.4) \quad \mathcal{C}_{::,k} = \mathbf{D}_{\mathbf{G}}^{(k)} \mathbf{F} \mathbf{D}_{\mathbf{H}}^{(k)},$$

with  $\mathbf{D}_{\mathbf{G}}^{(k)}, \mathbf{D}_{\mathbf{H}}^{(k)}$  as in (3.2). Injecting (4.4) in (4.3) and comparing with (1.1) we see that (4.3) is verified.  $\square$  This is a building block in computing the PT2D, as this corresponds to the subproblem of finding the remaining factors of the PT2D when  $\mathbf{A}$  and  $\mathbf{B}$  are known (see the next subsection for the details).

*Remark 4.9.* The ambiguities in Lemma 4.4 correspond exactly to ambiguities of PT2D in Definition 3.7 when  $\mathbf{A}$  and  $\mathbf{B}$  are fixed.

Lemma 4.8 shows that a PT2D tensor admits a Tucker-2 factorization  $\mathcal{C} \bullet_1 \mathbf{A} \bullet_2 \mathbf{B}$  with the core tensor structured as (4.1). In the next subsection we show how to use the Tucker-2 factorization to reduce dimension of the problem.



**4.4. Multilinear ranks in the PT2D.** Next, let us discuss how to reduce the PT2D decomposition problem from the case of rectangular factors  $\mathbf{A}$  and  $\mathbf{B}$  ( $I > R$  or  $J > S$ ) to the case when  $\mathbf{A}$  and  $\mathbf{B}$  are square (and invertible). For this, we use a standard strategy which performs Tucker compression first and relies on the following lemma.

LEMMA 4.10. *Let  $\mathcal{T}$  have an  $(R, S)$ -PT2D with  $R \leq I$  or  $S \leq J$ .*

1. *It holds that*

$$(4.5) \quad \text{rank}\{\mathbf{T}^{(1)}\} \leq R, \quad \text{rank}\{\mathbf{T}^{(2)}\} \leq S.$$

2. *Assume that  $\text{rank}\{\mathbf{T}^{(1)}\} = R, \text{rank}\{\mathbf{T}^{(2)}\} = S$ , and  $\mathbf{U} \in \mathbb{R}^{I \times R}, \mathbf{W} \in \mathbb{R}^{J \times S}$  be the bases of the corresponding column spaces (i.e.,  $\text{range}\{\mathbf{U}\} = \text{range}\{\mathbf{T}^{(1)}\}, \text{range}\{\mathbf{W}\} = \text{range}\{\mathbf{T}^{(2)}\}$ ). Then  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$  is an  $(R, S)$ -PT2D of  $\mathcal{T}$  if and only if the compressed tensor  $\mathcal{T}_c = \mathcal{T} \bullet_1 \mathbf{U}^\dagger \bullet_2 \mathbf{V}^\dagger$  has the PT2D*

$$(4.6) \quad \mathcal{T}_c = \text{PT2D}(\mathbf{A}_c, \mathbf{B}_c, \mathbf{F}, \mathbf{G}, \mathbf{H});$$

*Also, in (4.6), the factors of the compressed and full PT2D are linked as  $\mathbf{A} = \mathbf{U}\mathbf{A}_c, \mathbf{B} = \mathbf{V}\mathbf{B}_c$ .*

*Proof.* 1. From (1.1), we have that the  $\text{range}\{\mathcal{T}_{::,k}\} \subseteq \text{range}\{\mathbf{A}\}$  and  $\text{range}\{(\mathcal{T}_{::,k})^\top\} \subseteq \text{range}\{\mathbf{B}\}$ , which implies that  $\text{range}\{\mathbf{T}^{(1)}\} \subseteq \text{range}\{\mathbf{A}\}, \text{range}\{\mathbf{T}^{(2)}\} \subseteq \text{range}\{\mathbf{B}\}$ , hence the rank condition is satisfied.

2. If  $\text{rank}\{\mathbf{T}^{(1)}\} = R, \text{rank}\{\mathbf{T}^{(2)}\} = S$ , then it must hold that  $\text{range}\{\mathbf{T}^{(1)}\} = \text{range}\{\mathbf{A}\}$  and  $\text{range}\{\mathbf{T}^{(2)}\} = \text{range}\{\mathbf{B}\}$ , for any valid rank- $(R, S)$  PT2D. In particular, there exist nonsingular  $\mathbf{A}_c \in \mathbb{R}^{R \times R}$  and  $\mathbf{B}_c \in \mathbb{R}^{S \times S}$ , such that  $\mathbf{A} = \mathbf{U}\mathbf{A}_c$  and  $\mathbf{B} = \mathbf{V}\mathbf{B}_c$ . Then by multiplying  $\mathcal{T}_{::,k}$  by pseudoinverses of  $\mathbf{U}$  and  $\mathbf{V}$ , we get that the slices of the compressed tensor  $\mathcal{T}_c = \mathcal{T} \bullet_1 \mathbf{U}^\dagger \bullet_2 \mathbf{V}^\dagger$  can be expressed as

$$(\mathbf{U}^\dagger)^\top \mathcal{T}_{::,k} (\mathbf{V}^\dagger)^\top = \mathbf{A}_c \mathbf{D}_G^{(k)} \mathbf{F} \mathbf{D}_H^{(k)} \mathbf{B}_c^\top,$$

hence  $\mathcal{T}_c$  has a PT2D (4.6). Vice versa, any PT2D of  $\mathcal{T}_c$  in (4.6) gives rise to the PT2D of  $\mathcal{T}$  since  $\mathcal{T} = \mathcal{T}_c \bullet_1 \mathbf{U} \bullet_2 \mathbf{V}$ .  $\square$

In the rest of the subsection, we will show when the first and second unfoldings of  $\mathcal{T}$  are guaranteed to have maximal possible rank. We first propose some remarks on the ranks of unfoldings of  $\mathcal{C}$ .

PROPOSITION 4.11. *Let  $\mathbf{F}$  satisfy Assumption 2. Then the following holds true.*

1. *If there exists  $j$  such that  $\text{rank}\{\mathbf{G} \odot \mathbf{H}_{j,:}\} = R$ , then  $\text{rank}\{\mathbf{C}^{(1)}\} = R$ .*
2. *If there exists  $i$  such that  $\text{rank}\{\mathbf{G}_{i,:} \odot \mathbf{H}\} = S$ , then  $\text{rank}\{\mathbf{C}^{(2)}\} = S$ .*

*Proof.* See Subsection 7.3 for the proofs.  $\square$

Next, we can formulate the following simple corollary for the multilinear ranks of  $\mathcal{T}$ .

COROLLARY 4.12. *If both conditions of Proposition 4.11 are satisfied together with Assumption 1, then*

$$\text{rank}\{\mathbf{T}^{(1)}\} = R, \quad \text{rank}\{\mathbf{T}^{(2)}\} = S.$$

*Proof.* Follows from the fact that  $\mathbf{T}^{(1)} = \mathbf{A}\mathbf{C}^{(1)}, \mathbf{T}^{(2)} = \mathbf{B}\mathbf{C}^{(2)}$ , where both matrices in the factorizations have rank  $R$  and  $S$  respectively by Proposition 4.11 together with Assumption 1.  $\square$

Finally, we remark that the conditions of Proposition 4.11 are, in fact, satisfied under Assumption 5 or Assumption 6, as shown by the following corollary.

COROLLARY 4.13. *Under Assumption 2 and either of Assumption 5 or Assumption 6, we have*

$$\text{rank}\{\mathbf{C}^{(1)}\} = R, \quad \text{rank}\{\mathbf{C}^{(2)}\} = S.$$

*Proof.* See Subsection 7.3 for the proof.  $\square$

**4.5. Collinearities in  $\mathbf{F}$  and nonuniqueness.** In this section, we explain why Assumption 7 is necessary for uniqueness of the PT2D and why additional indeterminacies can arise in case Assumption 7 is not satisfied. This subsection can be skipped if the reader is not interested in nonunique cases. We start from a simple example to build an intuition.

*Example 4.14.* Let  $\text{rank}\{\mathbf{F}\} = 1$ , i.e., all the rows are collinear and the columns as well. Then thanks to [Lemma 3.6](#), we can assume, without loss of generality that  $\mathbf{F} = \mathbf{1}\mathbf{1}^\top$ . Then for the tensor  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$  decomposition of its slices (from (4.2) or (1.1)) reads simply

$$\mathcal{T}_{::,k} = \mathbf{A}(\mathbf{g}_k \mathbf{h}_k^\top) \mathbf{B}^\top.$$

Let us take any nonsingular matrices  $\mathbf{Y} \in \mathbb{R}^{R \times R}$ ,  $\mathbf{Z} \in \mathbb{R}^{S \times S}$ . Then if we define

$$\tilde{\mathbf{A}} = \mathbf{A}\mathbf{Y}, \quad \tilde{\mathbf{B}} = \mathbf{B}\mathbf{Z}, \quad \tilde{\mathbf{g}}_k = \mathbf{Y}^{-1}\mathbf{g}_k, \quad \tilde{\mathbf{h}}_k = \mathbf{Z}^{-1}\mathbf{h}_k,$$

we get that

$$\mathcal{T}_{::,k} = \tilde{\mathbf{A}}(\tilde{\mathbf{g}}_k \tilde{\mathbf{h}}_k^\top) \tilde{\mathbf{B}}^\top,$$

hence  $\mathcal{T} = \text{PT2D}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \mathbf{F}, \tilde{\mathbf{G}}, \tilde{\mathbf{H}})$  with  $\tilde{\mathbf{G}}$  and  $\tilde{\mathbf{H}}$  defined from  $\tilde{\mathbf{g}}_k, \tilde{\mathbf{h}}_k^\top$ .

[Example 4.14](#) means that for the rank-1  $\mathbf{F}$ , the factors  $\mathbf{A}$  and  $\mathbf{B}$  can be recovered uniquely in the PT2D only up to the subspaces spanned by their columns. In the general case ( $\text{krank}\{\mathbf{F}\} = 1$  or  $\text{krank}\{\mathbf{F}^\top\} = 1$ ) we will show that a similar statement can be made for blocks of columns of  $\mathbf{A}$  and  $\mathbf{B}$ .

Let  $\text{krank}\{\mathbf{F}\} = 1$ . Then, without loss of generality, we can assume that the columns of the matrix  $\mathbf{F}$  consists of blocks of repeating columns of sizes  $n_1, \dots, n_{S'}$ , i.e

$$\mathbf{F} = [\underbrace{\mathbf{f}_1 \cdots \mathbf{f}_1}_{n_1 \text{ times}} \cdots \underbrace{\mathbf{f}_{S'} \cdots \mathbf{f}_{S'}}_{n_{S'} \text{ times}}]$$

with  $n_1 + \cdots + n_{S'} = R$  and all  $\mathbf{f}_k$  are non-collinear ( $\text{krank}\{[\mathbf{f}_1 \quad \mathbf{f}_2 \quad \cdots \quad \mathbf{f}_{S'}]\} \geq 2$ ). Indeed, we can always reorder the rows to bring the collinear columns together and rescale the columns of  $\mathbf{F}$  such the collinear columns are represented by the same vector. This can be summarized in the following lemma.

**LEMMA 4.15.** *In PT2D with  $\mathbf{F}$  satisfying [Assumption 2](#) (and  $\text{rank}\{\mathbf{F}\} > 1$ ), we can assume that there exists  $\mathbf{F}' \in \mathbb{R}^{R' \times S'}$ ,  $R' \leq R, S' \leq S$ , and  $\text{krank}\{\mathbf{F}'\} \geq 2$ ,  $\text{krank}\{(\mathbf{F}')^\top\} \geq 2$ , so that  $\mathbf{F}$  can be represented<sup>6</sup> as*

$$(4.7) \quad \mathbf{F} = \begin{bmatrix} \mathbf{1}_{m_1} & & & \\ & \mathbf{1}_{m_2} & & \\ & & \ddots & \\ & & & \mathbf{1}_{m_{R'}} \end{bmatrix} \mathbf{F}' \begin{bmatrix} \mathbf{1}_{n_1}^\top & & & \\ & \mathbf{1}_{n_2}^\top & & \\ & & \ddots & \\ & & & \mathbf{1}_{n_{S'}}^\top \end{bmatrix}$$

where  $m_1 + \cdots + m_{R'} = R$  and  $n_1 + \cdots + n_{S'} = S$ .

Then for  $\mathbf{F}$  in a such form, we can describe the ambiguities that arise from the collinearities.

**PROPOSITION 4.16.** *Let  $\mathbf{F}$  be as in [Lemma 4.15](#), and  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$ . Then  $\mathcal{T}$  admits the family of following alternative PT2D (in addition to those that can be obtained in [Lemma 3.6](#)):*

$$\mathcal{T} = \text{PT2D}(\mathbf{A}\mathbf{Y}, \mathbf{B}\mathbf{Z}, \mathbf{F}, \mathbf{Y}^{-1}\mathbf{G}, \mathbf{Z}^{-1}\mathbf{H}),$$

where  $\mathbf{Y}$  and  $\mathbf{Z}$  are block-diagonal matrices

$$\mathbf{Y} = \text{Diag}\{\mathbf{Y}_1, \dots, \mathbf{Y}_{R'}\} = \begin{bmatrix} \mathbf{Y}_1 & & & \\ & \mathbf{Y}_2 & & \\ & & \ddots & \\ & & & \mathbf{Y}_{R'} \end{bmatrix}, \quad \mathbf{Z} = \text{Diag}\{\mathbf{Z}_1, \dots, \mathbf{Z}_{S'}\} = \begin{bmatrix} \mathbf{Z}_1 & & & \\ & \mathbf{Z}_2 & & \\ & & \ddots & \\ & & & \mathbf{Z}_{S'} \end{bmatrix},$$

with  $\mathbf{Y}_k \in \mathbb{R}^{m_k \times m_k}$  and  $\mathbf{Z}_k \in \mathbb{R}^{n_k \times n_k}$  are all invertible.

*Proof.* The proof is analogous to the derivations in [Example 4.14](#). □

<sup>6</sup>i.e., any  $\mathbf{F}$  satisfying [Assumption 2](#) can be brought to the form (4.7) by permutations and rescalings of columns and rows.

**5. Lifting approach to PT2D: the non-symmetric case.** In the next two subsections, we describe the main tool used in this paper, and present the main results for the non-symmetric PT2D. Our goal will be to determine  $\mathbf{A}$  and  $\mathbf{B}$ , since when these matrices are known, then we can find the other PT2D factors thanks to the results of the previous subsection. In the rest of the section, in view of [subsection 4.4](#), we assume that  $I = R$  and  $J = S$ , hence the matrices  $\mathbf{A}$  and  $\mathbf{B}$  are square invertible.

### 5.1. Lifting and the structured matrix: basic properties.

DEFINITION 5.1. For  $\mathcal{T} \in \mathbb{R}^{R \times S \times K}$ , we define the following  $(RS)^2 \times K$  structured matrix:

$$\Phi(\mathcal{T}) = [\text{vec}\{\mathcal{T}_{::,1}\} \boxtimes \text{vec}\{\mathcal{T}_{::,1}\} \quad \cdots \quad \text{vec}\{\mathcal{T}_{::,K}\} \boxtimes \text{vec}\{\mathcal{T}_{::,K}\}] \in \mathbb{R}^{(RS)^2 \times K},$$

or equivalently,  $\Phi(\mathcal{T})$  can be expressed as  $\Phi(\mathcal{T}) = (\mathbf{T}^{(3)})^\top \odot (\mathbf{T}^{(3)})^\top$ .

Remark 5.2. Note that the columns of  $\Phi(\mathcal{T})$ , in fact, correspond 4D tensors  $\{\mathcal{T}_{::,k} \otimes \mathcal{T}_{::,k}\}_{k=1}^K \subset \mathbb{R}^{R \times S \times R \times S}$  reshaped in a special way. This implies that  $\text{range}\{\Phi(\mathcal{T})\}$  consists of vectorized symmetric  $RS \times RS$  matrices, hence  $\text{range}\{\Phi(\mathcal{T})\} \subset \text{vec}\{S^2(\mathbb{R}^{RS})\}$  for any  $\mathcal{T}$ .

Now let us look at the properties of  $\Phi(\mathcal{T})$  when  $\mathcal{T}$  admits a PT2D, and particularly at the the left nullspace of  $\Phi(\mathcal{T})$ . Due to [Remark 5.2](#) only the symmetric part of the left kernel is informative. To this end, we will consider the symmetrization of the left nullspace  $\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})$ , which is, by [Remark 5.2](#), equal to

$$\sigma_2(\text{lker}\{\Phi(\mathcal{T})\}) = \text{lker}\{\Phi(\mathcal{T})\} \cap \text{vec}\{S^2(\mathbb{R}^{RS})\}.$$

Then the following key lemma summarizes the properties of rank and symmetric left nullspace properties for  $\Phi(\mathcal{T})$  and also explains the importance of [Assumption 5](#).

LEMMA 5.3. 1. For  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) \in \mathbb{R}^{R \times S \times K}$  with  $\mathbf{A} \in \mathbb{R}^{R \times R}$ ,  $\mathbf{B} \in \mathbb{R}^{S \times S}$  we have

$$(5.1) \quad \text{rank}\{\Phi(\mathcal{T})\} \leq \binom{R+1}{2} \binom{S+1}{2}.$$

2. If all [Assumptions 1, 2 and 5](#) are satisfied, then equality is achieved in (5.1). In this case,  $\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})$  spanned by  $M = \binom{R}{2} \binom{S}{2}$  vectors

$$\sigma_2(\text{lker}\{\Phi(\mathcal{T})\}) = \text{Span}\{\mathbf{p}_1, \dots, \mathbf{p}_M\}.$$

3. If one of [Assumptions 2 and 5](#) is not satisfied, we have a strict inequality in (5.1).

Proof. The proof is presented later in [Section 9](#). □

**5.2. The main result for non-symmetric case.** The main result relies on [Lemma 5.3](#) and needs an extra bit of notation. Let  $\pi : \text{vec}\{\mathbb{R}^{R \times S \times R \times S}\} \rightarrow \mathbb{R}^{R^2 \times S^2}$  be the permutation-symmetrization-matricization map that is defined on rank-one tensors as follows:

$$(5.2) \quad \pi : \text{vec}\{\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{y} \otimes \mathbf{z}\} \mapsto \frac{1}{2}(\mathbf{a} \boxtimes \mathbf{y} + \mathbf{y} \boxtimes \mathbf{a}) \cdot \frac{1}{2}(\mathbf{b} \boxtimes \mathbf{z} + \mathbf{z} \boxtimes \mathbf{b})^\top, \quad \text{for all } \mathbf{y}, \mathbf{a} \in \mathbb{R}^R, \mathbf{z}, \mathbf{b} \in \mathbb{R}^S,$$

that is  $\pi$  permutes modes 2 and 3 of the 4-way tensor, performs symmetrization along the modes of the same size, and flattens the tensor into an  $R^2 \times S^2$  matrix. Then the key result for the nonsymmetric PT2D can be formulated as follows.

THEOREM 5.4. Let  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) \in \mathbb{R}^{R \times S \times K}$ , whose factors satisfy [Assumptions 1, 2 and 5](#) and let  $\mathbf{P}_k = \pi(\mathbf{p}_k)$  where  $\mathbf{p}_k$  be any basis of  $\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})$  (as in [Lemma 5.3](#)).

1. Let  $\mathbf{P}_A = [\mathbf{P}_1 \quad \cdots \quad \mathbf{P}_M] \in \mathbb{R}^{R^2 \times S^2 K}$ . Then  $\dim(\sigma_2(\text{lker}\{\mathbf{P}_A\})) = R$  if and only if  $\text{krank}\{\mathbf{F}^\top\} \geq 2$ .
2. In that case

$$\sigma_2(\text{lker}\{\mathbf{P}_A\}) = \text{Span}\{\mathbf{a}_1 \boxtimes \mathbf{a}_1, \dots, \mathbf{a}_R \boxtimes \mathbf{a}_R\} = \text{range}\{\mathbf{A} \odot \mathbf{A}\},$$

where  $\mathbf{a}_k$  are the columns of  $\mathbf{A}$ .

2. Let  $\mathbf{P}_B = [\mathbf{P}_1^\top \quad \cdots \quad \mathbf{P}_M^\top] \in \mathbb{R}^{S^2 \times R^2 K}$ . Then  $\dim(\sigma_2(\text{lker}\{\mathbf{P}_B\})) = S$  if and only if  $\text{krank}\{\mathbf{F}\} \geq 2$ . In that case

$$\sigma_2(\text{lker}\{\mathbf{P}_B\}) = \text{Span}\{\mathbf{b}_1 \boxtimes \mathbf{b}_1, \dots, \mathbf{b}_S \boxtimes \mathbf{b}_S\} = \text{range}\{\mathbf{B} \odot \mathbf{B}\},$$

where  $\mathbf{b}_k$  are the columns of  $\mathbf{B}$ .

*Proof.* The proof will be split into several lemmas and presented later in a [Section 9](#).  $\square$

Then [Theorem 5.4](#) implies the following improved result on identifiability of PT2D.

**PROPOSITION 5.5.** *Let  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) \in \mathbb{R}^{I \times J \times K}$ , whose factors satisfy [Assumptions 1, 2, 5 and 7](#). Then the  $(R, S)$ -PT2D of  $\mathcal{T}$  is essentially unique. Also,  $(R, S)$  are minimal ranks (i.e., there does not exist  $(R', S')$ -PT2D with either  $R' < R$  or  $S' < S$ ).*

*Proof.* First, note that by [Assumptions 1, 2 and 5](#) we have that  $\text{rank}\{\mathbf{T}^{(1)}\} = R$  and  $\text{rank}\{\mathbf{T}^{(2)}\} = S$ , due to [Lemma 4.10](#). This implies that there is no  $(R', S')$ -PT2D with either  $R' < R$  or  $S' < S$  (otherwise we would have  $\text{rank}\{\mathbf{T}^{(1)}\} = R' < R$  or  $\text{rank}\{\mathbf{T}^{(2)}\} = S' < S$ , a contradiction). Next, due to [Lemma 4.10](#) we can assume, without loss of generality that  $I = R$  and  $J = S$ . If  $I > R$  and  $J > S$ , then by [Assumptions 1, 2 and 5](#) we have  $\text{rank}\{\mathbf{T}^{(1)}\} = R$  and  $\text{rank}\{\mathbf{T}^{(2)}\} = S$ , thus by [Lemma 4.10](#), the  $(R, S)$ -PT2D of  $\mathcal{T}$  is unique if and only if the  $(R, S)$ -PT2D of the compressed tensor  $\mathcal{T}_c \in \mathbb{R}^{R \times S \times K}$  is unique.

Now, let us assume  $I = R, J = S$ . Under [Assumptions 1, 2 and 5](#), by [Lemma 5.3](#) we have  $\text{rank}\{\Phi(\mathcal{T})\} = \binom{R+1}{2} \binom{S+1}{2}$ . If [Assumption 7](#) is also satisfied, we have  $\dim(\sigma_2(\text{lker}\{\mathbf{P}_A\})) = R$  and  $\dim(\sigma_2(\text{lker}\{\mathbf{P}_B\})) = S$  by [Theorem 5.4](#).

Let  $\mathbf{q}_{A,1}, \dots, \mathbf{q}_{A,R} \in \mathbb{R}^{R^2}$  be some basis of  $\sigma_2(\text{lker}\{\mathbf{P}_A\})$ . Then, by [Theorem 5.4](#) there is a nonsingular (change of basis) matrix  $\mathbf{W}_A \in \mathbb{R}^{R \times R}$

$$[\mathbf{q}_{A,1} \quad \dots \quad \mathbf{q}_{A,R}] = (\mathbf{A} \odot \mathbf{A})(\mathbf{W}_A)^\top,$$

This implies that the tensor  $\mathcal{Q}_A \in \mathbb{R}^{R \times R \times R}$  built as  $(\mathcal{Q}_A)_{::,k} = \text{matr}_{R,R}\{\mathbf{q}_{A,k}\}$  admits a CPD

$$\mathcal{Q}_A = [\mathbf{A}, \mathbf{A}, \mathbf{W}_A].$$

Note that, by our assumptions,  $\text{krank}\{\mathbf{A}\} = \text{krank}\{\mathbf{W}_A\} = R$ , and therefore the Kruskal's uniqueness condition is satisfied [16], and the matrix  $\mathbf{A}$  is unique up to permutation and scaling of columns.

Similarly, we can construct  $\mathcal{Q}_B \in \mathbb{R}^{S \times S \times S}$  built as  $(\mathcal{Q}_B)_{::,k} = \text{matr}_{S,S}\{\mathbf{q}_{B,k}\}$  (with  $\{\mathbf{q}_{B,k}\}_{k=1}^S$  be any basis of  $\sigma_2(\text{lker}\{\mathbf{P}_B\})$ ). For this tensor, we have that, for some nonsingular  $\mathbf{W}_B \in \mathbb{R}^{S \times S}$ ,

$$\mathcal{Q}_B = [\mathbf{B}, \mathbf{B}, \mathbf{W}_B],$$

whose CPD is unique by Kruskal's theorem and  $\mathbf{B}$  is unique up to permutation and scaling of columns. Thus, we can find  $\mathcal{C} = \mathcal{T} \bullet_1 \mathbf{A}^{-1} \bullet \mathbf{B}^{-1}$  which must be of the form (4.1). The factors  $\mathbf{F}, \mathbf{G}, \mathbf{H}$  can be recovered from  $\mathcal{C}$  up to scaling indeterminacies, thanks to [Proposition 4.7](#).

Finally, note that for any alternative  $(R, S)$ -PT2D,  $\mathcal{T} = \text{PT2D}(\tilde{\mathbf{A}}, \tilde{\mathbf{B}}, \tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \tilde{\mathbf{H}})$  we have that: we have

- [Assumption 1](#) for  $(\tilde{\mathbf{A}}, \tilde{\mathbf{B}})$  must be satisfied since  $\text{rank}\{\mathbf{T}^{(1)}\} = R$  and  $\text{rank}\{\mathbf{T}^{(2)}\} = S$ .
- [Assumptions 2 and 5](#) must be satisfied as well from statement 3) of [Lemma 5.3](#);
- and [Assumption 7](#) must be satisfied due to “if and only if” statements in and [Theorem 5.4](#).

Therefore, (essential) uniqueness of the factors  $\mathbf{A}$  and  $\mathbf{B}$  in the CPDs of  $\mathcal{Q}_A$  and  $\mathcal{Q}_B$  implies uniqueness of  $\mathbf{A}$  and  $\mathbf{B}$  in  $(R, S)$ -PT2D.  $\square$

**5.3. Algorithm for nonsymmetric case.** The proof of [Proposition 5.5](#) and [Theorem 5.4](#) suggest a way to compute the (unique under [Assumptions 1, 2, 5 and 7](#)) PT2D of a given tensor, and we summarize it in [Algorithm 5.1](#). We also show that the assumptions on factors ([Assumptions 1, 2, 5 and 7](#)) can be replaced by the assumptions on the tensor  $\mathcal{T}$  which can be checked *a posteriori*.

*Remark 5.6.* [Algorithm 5.1](#) can be also applied in the approximation scenario, where all the exact decomposition steps are replaced by approximations. For example, the nullspace computations can be performed by using the SVD, and the CPD can be replaced by CP approximation.

*Remark 5.7.* The algorithm can be modified to handle the case when [Assumption 5](#) is not satisfied. In this case, the nullspaces  $\sigma_2(\text{lker}\{\mathbf{P}_A\})$ ,  $\sigma_2(\text{lker}\{\mathbf{P}_B\})$  will become larger and instead of the CPD of  $\hat{\mathcal{Q}}_A$ ,  $\hat{\mathcal{Q}}_B$  we would need to perform the symmetric joint block diagonalization. See [Subsection 9.3](#) for more details.

*Remark 5.8* (From assumptions on factors to assumptions on the tensor). The proof of [Proposition 5.5](#) shows that [Assumptions 1, 2, 5 and 7](#) hold if and only if  $\text{rank}\{\mathbf{T}^{(1)}\} = R$  and  $\text{rank}\{\mathbf{T}^{(2)}\} = S$  and  $\text{rank}\{\Phi(\mathcal{T}_c)\} = \binom{R+1}{2} \binom{S+1}{2}$ , and  $\mathcal{T}$  admits an  $(R, S)$ -PT2D. Therefore [Algorithm 5.1](#) returns the unique decomposition if the “implicit” conditions on  $\mathcal{T}$  are satisfied.

**Algorithm 5.1** Algebraic decomposition algorithm for nonsymmetric ParaTuck-2

- 
- 1: **Input:**  $\mathcal{T} \in \mathbb{R}^{I \times J \times K}$  ranks  $(R, S)$
  - 2: Compute Tucker-2 decomposition:  $\mathcal{T} = \mathcal{T}_c \bullet_1 \mathbf{U} \bullet_2 \mathbf{V}$  with  $\mathbf{U} \in \mathbb{R}^{I \times R}$ ,  $\mathbf{V} \in \mathbb{R}^{J \times S}$
  - 3: Find the  $M = \binom{R}{2} \binom{S}{2}$  vectors  $\mathbf{p}_1, \dots, \mathbf{p}_M$  vector in symmetric left kernel of  $\sigma_2(\text{lker}\{\Phi(\mathcal{T}_c)\})$
  - 4: Find the projections  $\mathbf{P}_k = \pi(\mathbf{p}_k)$ , build the matrices  $\mathbf{P}_A$  and  $\mathbf{P}_B$ .
  - 5: Compute bases  $\{\mathbf{q}_{A,1}, \dots, \mathbf{q}_{A,R}\}$  and  $\{\mathbf{q}_{B,1}, \dots, \mathbf{q}_{B,S}\}$  of the symmetric left kernels  $\sigma_2(\text{lker}\{\mathbf{P}_A\})$  and  $\sigma_2(\text{lker}\{\mathbf{P}_B\})$ , respectively.
  - 6: Stack these matrices in tensors  $\hat{\mathbf{Q}}_A$  and  $\hat{\mathbf{Q}}_B$  and compute their CPD  
 $\hat{\mathbf{Q}}_A = [\hat{\mathbf{A}}_c, \hat{\mathbf{A}}_c, \mathbf{W}_A]$  and  $\hat{\mathbf{Q}}_B = [\hat{\mathbf{B}}_c, \hat{\mathbf{B}}_c, \mathbf{W}_B]$ .
  - 7: Set  $\hat{\mathbf{A}} = \mathbf{U} \hat{\mathbf{A}}_c$ ,  $\hat{\mathbf{B}} = \mathbf{V} \hat{\mathbf{B}}_c$ .
  - 8: Find the core tensor  $\hat{\mathcal{C}} = \mathcal{T}_c \bullet_1 \hat{\mathbf{A}}_c^\dagger \bullet_2 \hat{\mathbf{B}}_c^\dagger$ .
  - 9: Determine the factors  $\hat{\mathbf{F}}$ ,  $\hat{\mathbf{G}}$  and  $\hat{\mathbf{H}}$  using the methods from [Subsection 4.2](#) (namely, [Proposition 4.7](#)).
  - 10: **return**  $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{F}}, \hat{\mathbf{G}}, \hat{\mathbf{H}}$ , such that  $\mathcal{T} = \text{PT2D}(\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{F}}, \hat{\mathbf{G}}, \hat{\mathbf{H}})$
- 

*SVD- and EVD-based specialization of the algorithm.* Let us comment on how different steps of [Algorithm 5.1](#) can be performed, by using only standard linear algebra operations, in the view of performing possibly the approximate ParaTuck-2.

- Step 2 can be computed by the HOSVD: we take  $\mathbf{U} \in \mathbb{R}^{I \times R}$ ,  $\mathbf{V} \in \mathbb{R}^{J \times S}$  the matrices of the leading  $R, S$  left singular vectors of  $\mathbf{T}^{(1)}$  and  $\mathbf{T}^{(2)}$ , respectively. In this case, the core (compressed) tensor  $\mathcal{T}_c \in \mathbb{R}^{R \times S \times K}$  is found as  $\mathcal{T}_c = \mathcal{T} \bullet_1 \mathbf{U}^\top \bullet_2 \mathbf{V}^\top$
- In particular,  $R$  and  $S$  can be estimated at the step 2 in case they are not given *a priori* (i.e., they can be estimated from the singular values of  $\mathbf{T}^{(1)}$  and  $\mathbf{T}^{(2)}$ ).
- Step 3 can be performed by using the SVD of the matrix  $\sigma_2(\text{lker}\{\Phi(\mathcal{T}_c)\})$ . There are two options.
  - The simplest is to compute the last  $(RS)^2 - \binom{R+1}{2} \binom{S+1}{2}$  left singular vectors of  $\sigma_2(\text{lker}\{\Phi(\mathcal{T}_c)\})$ , symmetrize them with  $\sigma_2(\cdot)$ , and extract a basis of the symmetrized singular vectors.
  - Alternatively, we can form the matrix  $\Phi'(\mathcal{T}_c) \in \mathbb{R}^{\binom{RS+1}{2} \times K}$  by removing the repeating rows (which can only repeat twice) and multiplying the rows which had a repetition by a factor of  $\sqrt{2}$ , computing the last  $M = \binom{R}{2} \binom{S}{2}$  left singular vectors of  $\Phi'(\mathcal{T}_c)$  (denoted as  $\mathbf{p}'_1, \dots, \mathbf{p}'_M$ ) and expanding  $\mathbf{p}'_k$  to  $\mathbf{p}_k$  by repeating the corresponding elements and dividing them by  $\sqrt{2}$ .
- In steps 5 and 6: we can avoid extra symmetrization in  $\pi$ . In fact, let  $\pi'$  be just the map that maps  $\text{vec}\{\text{vec}\{\mathbf{a} \otimes \mathbf{b} \otimes \mathbf{y} \otimes \mathbf{z}\} \mapsto (\mathbf{y} \boxtimes \mathbf{a}) \cdot \frac{1}{2}(\mathbf{z} \boxtimes \mathbf{b})^\top\}$ . Then for any  $\mathbf{p} \in \text{vec}\{S^2(\mathbb{R}^{RS})\}$

$$\text{range}\{\pi(\mathbf{p})\} = \sigma_2(\text{range}\{\pi'(\mathbf{p})\}) \text{ and } \text{range}\{\pi(\mathbf{p})^\top\} = \sigma_2(\text{range}\{\pi'(\mathbf{p})^\top\}),$$

hence we can replace  $\pi$  with  $\pi'$  and perform the symmetrization at the moment of computing  $\sigma_2(\text{lker}\{\mathbf{P}_A\})$  and  $\sigma_2(\text{lker}\{\mathbf{P}_B\})$ .

- For  $\hat{\mathbf{Q}}_A$ ,  $\hat{\mathbf{Q}}_B$  so that we can find the CPD (approximation) by a joint eigenvalue decomposition of several matrices. Indeed, we have for any index  $k$

$$(\hat{\mathbf{Q}}_A)_{k,:,:} = \hat{\mathbf{A}}_c \text{Diag}\{(\hat{\mathbf{A}}_c)_{k,:}\} \mathbf{W}_A^\top,$$

and therefore

$$\mathbf{Y}_{k,\ell} = (\hat{\mathbf{Q}}_A)_{k,:,:}^{-1} (\hat{\mathbf{Q}}_A)_{\ell,:,:} = \mathbf{W}_A^{-\top} \mathbf{D}_{k,\ell} \mathbf{W}_A^\top$$

with  $\mathbf{D}_{k,\ell} = \text{Diag}\{(\hat{\mathbf{A}}_c)_{k,:}\}^{-1} \text{Diag}\{(\hat{\mathbf{A}}_c)_{\ell,:}\}$ . Therefore, we can fix  $k$  (say, to  $k = 1$ , which will choose a normalization  $(\hat{\mathbf{A}}_c)_{1,:} = 1$ ), and take  $\ell = 2, \dots, R$ , to perform joint eigenvalue decomposition of  $\mathbf{Y}_{1,2}, \dots, \mathbf{Y}_{1,R}$  (for example, by the matrix pencil method). This will give the remaining  $(\hat{\mathbf{A}}_c)_{\ell,:}$ . To avoid ill-conditioning, we can premultiply the first and second modes of the tensor  $\hat{\mathbf{Q}}_A$  by a random rotation.

**6. Symmetric case (DEDICOM).** As in the non-symmetric case, we apply the lifting approach and the algorithm relies on the finding the left nullspace of  $\Phi(\mathcal{T})$ . However, there are important differences with respect to the nonsymmetric case.

**6.1. Properties of the structured matrix.** Note that the column span of  $\Phi(\mathcal{T})$  has additional symmetry structure, in case when all slices  $\mathcal{T}_{:, :, k}$  are symmetric matrices. Recall the definition of the subspace  $\text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\} \subset \mathbb{R}^{R \times R \times R \times R}$ , which is

$$\text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\} = \text{Span}\{\text{vec}\{\mathbf{X}\} \boxtimes \text{vec}\{\mathbf{X}\}, \mathbf{X} \in S^2(\mathbb{R}^R)\}.$$

In fact,  $\text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\}$  is the space spanned by the vectorizations of tensors  $\mathcal{Y} \in \mathbb{R}^{R \times R \times R \times R}$  satisfying the following sets of symmetries

$$\mathcal{Y}_{ijkl} = \mathcal{Y}_{jikl} = \mathcal{Y}_{ijlk} = \mathcal{Y}_{klij}.$$

This implies the following remark.

*Remark 6.1.* For  $\mathcal{T} = \text{DEDICOM}(\mathbf{A}, \mathbf{F}, \mathbf{G}) \in \mathbb{R}^{R \times R \times K}$ , the column space of  $\Phi(\mathcal{T})$  is a subspace of the space of partially symmetric tensors defined above.

$$\text{range}\{\{\Phi(\mathcal{T})\}\} \subseteq \text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\}$$

We will introduce the following useful definition.

**DEFINITION 6.2.** Let  $\sigma_{(2,2)} : \mathbb{R}^{R^4} \rightarrow \mathbb{R}^{R^4}$  denote the orthogonal projection on  $\text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\}$ , i.e. it is the map that acts on rank-one tensors as:

$$\sigma_{(2,2)} : \mathbf{z} \boxtimes \mathbf{y} \boxtimes \mathbf{b} \boxtimes \mathbf{a} \mapsto \sigma(\sigma(\mathbf{z} \boxtimes \mathbf{y}) \boxtimes \sigma(\mathbf{b} \boxtimes \mathbf{a})), \quad \text{for all } \mathbf{z}, \mathbf{y}, \mathbf{b}, \mathbf{a} \in \mathbb{R}^R.$$

Then for any matrix  $\mathbf{Z} \in \mathbb{R}^{R^4 \times K}$ , we define the symmetrized left kernel as  $\sigma_{(2,2)}(\text{lker}\{\mathbf{Z}\})$

Note that due to [Remark 6.1](#), we have (for any  $\mathcal{T} = \text{DEDICOM}(\mathbf{A}, \mathbf{F}, \mathbf{G})$ )

$$\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\}) = \text{lker}\{\Phi(\mathcal{T})\} \cap \text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\}$$

Armed with the definition of the symmetrized kernel, we can formulate the following analogue of [Lemma 5.3](#).

**LEMMA 6.3.** 1. For  $\mathcal{T} = \text{DEDICOM}(\mathbf{A}, \mathbf{F}, \mathbf{G}) \in \mathbb{R}^{R \times S \times K}$  with  $\mathbf{A} \in \mathbb{R}^{R \times R}$  and  $\mathbf{F}$  symmetric we have

$$(6.1) \quad \text{rank}\{\Phi(\mathcal{T})\} \leq \binom{R+3}{4},$$

2. If all [Assumptions 1, 2 and 6](#) are satisfied, then the equality in (6.1) is achieved. In this case,  $\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\})$  spanned by  $N = \frac{R^2(R-1)(R+1)}{12}$  vectors

$$\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\}) = \text{Span}\{\mathbf{p}_1, \dots, \mathbf{p}_N\}.$$

3. If either of [Assumptions 2 and 6](#) is not satisfied, we have a strict inequality in (6.1).

*Proof.* The proof is given in [Section 10](#). □

**6.2. Main result for the symmetric case.** Recall that for any matrix  $\mathbf{Z} \in \mathbb{R}^{R^3 \times N}$ , the symmetrization of the left kernel is defined as  $\sigma_3(\text{lker}\{\mathbf{Z}\})$ .

**THEOREM 6.4.** Let  $\mathcal{T} = \text{DEDICOM}(\mathbf{A}, \mathbf{F}, \mathbf{G}) \in \mathbb{R}^{R \times S \times K}$ , whose factors satisfy [Assumptions 1, 2 and 6](#) and let  $\mathbf{P}_k = \text{matr}_{S^3, S}\{\sigma_4(\mathbf{p}_n)\}$  where  $\{\mathbf{p}_n\}_{n=1}^N$  be any basis of  $\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\})$  (as in [Lemma 6.3](#)). Consider the matrix

$$(6.2) \quad \mathbf{P}_{\text{sym}} = [\mathbf{P}_1 \quad \dots \quad \mathbf{P}_N]$$

Then,  $\text{rank}\{\mathbf{P}_{\text{sym}}\} \leq \binom{R+2}{3} - R$  and the equality holds if and only if [Assumption 7](#) holds. In the latter case, we have

$$\sigma_3(\text{lker}\{\mathbf{P}_{\text{sym}}\}) = \text{Span}\{\mathbf{a}_1 \boxtimes \mathbf{a}_1 \boxtimes \mathbf{a}_1, \dots, \mathbf{a}_R \boxtimes \mathbf{a}_R \boxtimes \mathbf{a}_R\} = \text{range}\{\mathbf{A} \odot \mathbf{A} \odot \mathbf{A}\}.$$

*Proof.* The proof will be split into several lemmas and presented later in [Section 10](#). □

Then [Theorem 6.4](#) implies the following improved result on identifiability of PT2D.



PROPOSITION 6.5. Let  $\mathcal{T} = \text{DEDICOM}(\mathbf{A}, \mathbf{F}, \mathbf{G}) \in \mathbb{R}^{I \times I \times K}$ ,  $\mathbf{F} \in \mathbb{R}^{R \times R}$  symmetric, whose factors satisfy Assumptions 1, 2, 6 and 7. Then the  $R$ -DEDICOM of  $\mathcal{T}$  is essentially unique. In addition,  $R$  is the minimal possible rank for a DEDICOM of  $\mathcal{T}$ .

*Proof.* Again, due to Lemma 4.10 and Assumptions 1, 2 and 6, we have that  $\text{rank}\{\mathcal{T}^{(1)}\} = R$  which implies that there is no  $R'$ -DEDICOM for  $R' < R$ . Also, we can assume, without loss of generality that  $I = R$ . If  $I > R$ , then we can first perform Tucker compression on  $\mathcal{T}$  (see Lemma 4.10) and compute the PT2D on the compressed tensor, as in the proof of Proposition 6.5.

Under Assumptions 1, 2 and 6, by Lemma 5.3 we have  $\text{rank}\{\Phi(\mathcal{T})\} = \binom{R+3}{4}$ . If Assumption 7 is also satisfied, we have  $\dim(\sigma_3(\text{lker}\{\mathbf{P}_{sym}\})) = R$  by Theorem 6.4.

Let  $\mathbf{q}_1, \dots, \mathbf{q}_R \in \mathbb{R}^{R^3}$  be some basis of  $\sigma_3(\text{lker}\{\mathbf{P}_{sym}\})$ . Then, by Theorem 5.4 there is a nonsingular (change of basis) matrix  $\mathbf{W} \in \mathbb{R}^{R \times R}$

$$[\mathbf{q}_1 \quad \dots \quad \mathbf{q}_R] = (\mathbf{A} \odot \mathbf{A} \odot \mathbf{A})\mathbf{W}^\top,$$

Which implies that the tensor  $\mathcal{Q} \in \mathbb{R}^{R \times R \times R \times R}$  built as  $(\mathcal{Q})_{:, :, :, k} = \text{tens}_{R, R, R}\{\mathbf{q}_k\}$  has a CPD

$$\mathcal{Q} = [\mathbf{A}, \mathbf{A}, \mathbf{A}, \mathbf{W}].$$

Note that, by our assumptions,  $\text{krank}\{\mathbf{A}\} = \text{krank}\{\mathbf{W}\} = R$ . Thus the Kruskal's uniqueness conditions are satisfied and the matrix  $\mathbf{A}$  is unique up to permutation and scaling of columns. The factors  $\mathbf{F}, \mathbf{G}$  can be recovered from  $\mathcal{C}$  up to scaling indeterminacies, thanks to Proposition 4.7.

Finally, due to “if and only if” statements in Lemma 6.3 and Theorem 6.4, for any alternative  $(R)$ -DEDICOM,  $\mathcal{T} = \text{DEDICOM}(\tilde{\mathbf{A}}, \tilde{\mathbf{F}}, \tilde{\mathbf{G}})$ , all Assumptions 1, 2, 6 and 7 must be satisfied as well.  $\square$

**6.3. Algorithm for the symmetric case.** The proofs of Proposition 6.5 and Theorem 6.4 are also constructive. We summarize in Algorithm 6.1 the algorithms for computing the (unique)  $R$ -DEDICOM of a given tensor, under Assumptions 1, 2 and 7 and Assumption 6, respectively.

---

**Algorithm 6.1** Algebraic decomposition algorithm for DEDICOM

---

- 1: **Input:**  $\mathcal{T} \in \mathbb{R}^{I, I, K}$  rank  $(R)$
  - 2: Compute the symmetric Tucker-2 decomposition (approximation)  $\mathcal{T} = \mathcal{T}_c \bullet_1 \mathbf{U} \bullet_2 \mathbf{U}$  with  $\mathbf{U} \in \mathbb{R}^{I \times R}$
  - 3: Find the  $N$  vectors  $\mathbf{p}_1, \dots, \mathbf{p}_N$  in (approximate) symmetric left kernel  $\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T}_c)\})$
  - 4: Find the symmetrizations  $\mathbf{P}_k = \text{matr}_{R^3, R}\{\sigma_4(\mathbf{p}_k)\}$ , build  $\mathbf{P}_{sym}$  from  $\mathbf{P}_k$ .
  - 5: Compute a basis  $\{\mathbf{q}_1, \dots, \mathbf{q}_R\}$  of the (approximate) symmetric left kernel  $\sigma_3(\text{lker}\{\mathbf{P}_{sym}\})$ .
  - 6: Stack these matrices in tensor  $\hat{\mathcal{Q}}_{sym} \in \mathbb{R}^{R \times R \times R \times R}$  and compute its CPD (approximation)  
 $\hat{\mathcal{Q}}_{sym} = [\hat{\mathbf{A}}_c, \hat{\mathbf{A}}_c, \hat{\mathbf{A}}_c, \mathbf{W}_A]$ .
  - 7: Set  $\hat{\mathbf{A}} = \mathbf{U} \hat{\mathbf{A}}_c$  and find the core tensor  $\hat{\mathcal{C}} = \mathcal{T}_c \bullet_1 \hat{\mathbf{A}}_c^\dagger \bullet_2 \hat{\mathbf{A}}_c^\dagger$ .
  - 8: Determine the factors  $\hat{\mathbf{F}}, \hat{\mathbf{G}}$  using the methods from Subsection 4.2 (namely, Proposition 4.7).
  - 9: **return**  $\hat{\mathbf{A}}, \hat{\mathbf{F}}, \hat{\mathbf{G}}$ , such that  $\mathcal{T} = \text{DEDICOM}(\hat{\mathbf{A}}, \hat{\mathbf{F}}, \hat{\mathbf{G}})$
- 

As in the nonsymmetric algorithm, we make remarks on each of the individual steps of the algorithm. *SVD- and EVD- based specializations of the algorithm.*

- In step 2, the HOSVD again is used: we can compute  $\mathbf{U}^{I \times R}$  from singular vectors and also estimate  $R$  if needed.
- The symmetrized kernel  $\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T}_c)\})$  can be computed as in the nonsymmetric case.
- The symmetrization  $\sigma_4$  can be skipped and combined with computing  $\sigma_3(\text{lker}\{\mathbf{P}_{sym}\})$ .
- The CPD of  $\hat{\mathcal{Q}}_{sym}$  can be computed by joint eigenvalue decomposition of  $R \times R$  matrices

$$\mathbf{Y}_{k, \ell} = (\text{matr}_{R^2, R}\{(\hat{\mathcal{Q}}_A)_{k, :, :, :}\})^\dagger \text{matr}_{R^2, R}\{(\hat{\mathcal{Q}}_A)_{\ell, :, :, :}\}.$$

**7. Properties of the core tensor.** In this section, we review advanced properties of the core tensor that serve as the base for the proofs.

**7.1. Implicit equations for the core tensor and their determinantal representation.** We begin by recalling some equations that must be satisfied by a core tensor

LEMMA 7.1. *Let  $\mathcal{C} \in \mathbb{R}^{R \times S \times K}$  be a tensor of the form (4.1), i.e.,  $\mathcal{C} = \text{TP}(\mathbf{F}, \mathbf{G}, \mathbf{H})$ . Then for all  $1 \leq i, r \leq R$ ,  $1 \leq j, s \leq S$ ,  $1 \leq k, t \leq K$  it holds that*

$$(7.1) \quad \mathcal{C}_{ijk}\mathcal{C}_{rsk}\mathcal{C}_{ist}\mathcal{C}_{rjt} - \mathcal{C}_{isk}\mathcal{C}_{rjk}\mathcal{C}_{ijt}\mathcal{C}_{rst} = 0.$$

*Proof.* Follows from straightforward substitution, as

$$\mathcal{C}_{ijk}\mathcal{C}_{rsk}\mathcal{C}_{ist}\mathcal{C}_{rjt} = F_{ij}F_{rs}F_{is}F_{rj}G_{ik}G_{rk}G_{it}G_{rt}H_{jk}H_{sk}H_{jt}H_{st} = \mathcal{C}_{isk}\mathcal{C}_{rjk}\mathcal{C}_{ijt}\mathcal{C}_{rst} \quad \square$$

The equation (7.1) can be viewed as vanishing of a “generalized determinant” for  $2 \times 2 \times 2$  tensors (see (7.1)). Note that there are other polynomial equations for tensors satisfying (7.1), and describing them is a difficult task [10, end of §6]. However, for our purposes, equations (7.1) will be largely sufficient.

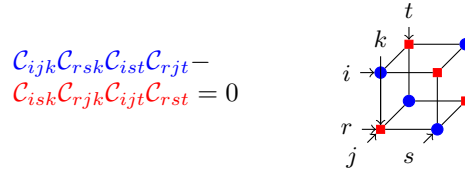


Fig. 7.1: Equations for core tensor via in terms of the elements of the  $2 \times 2 \times 2$  subtensor  $\mathcal{C}_{(i,r),(j,s),(k,t)}$ .

Note that the equation (7.1) can be also rewritten as

$$(7.2) \quad \det \begin{bmatrix} \mathcal{C}_{ijk}\mathcal{C}_{rsk} & \mathcal{C}_{ijt}\mathcal{C}_{rst} \\ \mathcal{C}_{isk}\mathcal{C}_{rjk} & \mathcal{C}_{ist}\mathcal{C}_{rjt} \end{bmatrix} = 0$$

Then equations (7.2) for all  $1 \leq k < t \leq K$  can be grouped to obtain the following result.

LEMMA 7.2. *Let  $\mathcal{C} \in \mathbb{R}^{R \times S \times K}$  be a tensor such that  $\mathcal{C} = \text{TP}(\mathbf{F}, \mathbf{G}, \mathbf{H})$ .*

1. *Then for all  $1 \leq i, r \leq R$ ,  $1 \leq j, s \leq S$ , the matrix*

$$\Psi^{(ijrs)} := \begin{bmatrix} \mathcal{C}_{ij1}\mathcal{C}_{rs1} & \cdots & \mathcal{C}_{ijK}\mathcal{C}_{rsK} \\ \mathcal{C}_{is1}\mathcal{C}_{rj1} & \cdots & \mathcal{C}_{isK}\mathcal{C}_{rjK} \end{bmatrix} \in \mathbb{R}^{2 \times K}$$

*must be at most rank 1.*

2. *In particular, the vector*

$$(7.3) \quad \begin{bmatrix} \theta_1^{(ijrs)} & \theta_2^{(ijrs)} \end{bmatrix} := \begin{bmatrix} -F_{is}F_{rj} & F_{ij}F_{rs} \end{bmatrix},$$

*lies in the left kernel of  $\Psi^{(ijrs)}$ , i.e.,*

$$(7.4) \quad \begin{bmatrix} \theta_1^{(ijrs)} & \theta_2^{(ijrs)} \end{bmatrix} \Psi^{(ijrs)} = 0.$$

*Proof.* 1. Let us fix  $1 \leq i, r \leq R$ ,  $1 \leq j, s \leq S$ . Then, thanks to (7.2), vanishing of all  $2 \times 2$  minors of  $\Psi^{(ijrs)}$  is equivalent to the inequality  $\text{rank}\{\Psi^{(ijrs)}\} \leq 1$ .

2. The straightforward computation leads to

$$(\Psi^{(ijrs)})_{:,k} = (G_{ik}G_{rk}H_{jk}H_{sk})\mathbf{z}^{(ijrs)}, \quad \text{with } \mathbf{z}^{(ijrs)} = \begin{bmatrix} F_{ij}F_{rs} \\ F_{is}F_{rj} \end{bmatrix},$$

or, equivalently

$$(7.5) \quad \Psi^{(ijrs)} = \mathbf{z}^{(ijrs)}(\mathbf{G}_{i,:} \odot \mathbf{G}_{r,:} \odot \mathbf{H}_{j,:} \odot \mathbf{H}_{s,:})$$

Note that the vector  $\begin{bmatrix} \theta_1^{(ijrs)} & \theta_2^{(ijrs)} \end{bmatrix}^\top$  is obviously orthogonal to  $\mathbf{z}^{(ijrs)}$  and the proof is complete.  $\square$

### 7.2. Factorizing the core tensor: proofs.

*Proof Lemma 4.5.* Indeed, the condition of lemma implies  $F_{ij} \neq 0$  and hence we can take  $\tilde{\mathbf{F}} = \mathcal{C}_{:, :, r}$  and  $\tilde{\mathbf{g}}_r = \mathbf{1}$  and  $\tilde{\mathbf{h}}_r = \mathbf{1}$ . Now, for fixed  $k$  let  $\mathbf{Z}_k \in \mathbb{R}^{R \times S}$  be defined by elementwise division  $(\mathbf{Z}_k)_{i,j} = \mathcal{C}_{i,j,r} / \tilde{F}_{ij}$ . Then we can find the  $k$ -th columns of  $\tilde{\mathbf{G}}$  and  $\tilde{\mathbf{H}}$  (i.e.,  $\tilde{\mathbf{g}}_k$  and  $\tilde{\mathbf{h}}_k$ ) from the rank-one factorization of  $\mathbf{Z}_k$ :

$$\tilde{\mathbf{g}}_k(\tilde{\mathbf{h}}_k)^\top = \mathbf{Z}_k.$$

It is easy to see that for such matrices we have  $\mathcal{C} = \text{TP}(\tilde{\mathbf{F}}, \tilde{\mathbf{G}}, \tilde{\mathbf{H}})$ , thus it gives a valid factorization of  $\mathcal{C}$ .  $\square$

*Proof of Proposition 4.7.* By the scaling ambiguities, we can take always take  $\mathbf{F}$  to have the first row and column consisting of all ones. Then we have that

$$(7.6) \quad \mathbf{F} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & * & \cdots & * \\ \vdots & \vdots & & \vdots \\ 1 & * & \cdots & * \end{bmatrix}.$$

We will show that we can complete the  $*$  elements uniquely.

If [Assumption 5](#) or [Assumption 6](#) is satisfied, neither of rows of  $\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}$  is identically zero. That is, for arbitrary  $1 < r \leq R$ ,  $1 < s \leq S$ ,

$$(\mathbf{G}_{1,:} \odot \mathbf{G}_{r,:} \odot \mathbf{H}_{1,:} \odot \mathbf{H}_{s,:}) \neq \mathbf{0}$$

and hence  $\text{rank}\{\Psi^{(11rs)}\} = 1$  in view of (7.5). Therefore, we can determine  $\theta_1^{(11rs)}, \theta_2^{(11rs)}$  in (7.4) up to a constant from  $\text{lker}\{\Psi^{(11rs)}\}$ , and thus

$$F_{rs} = \frac{F_{rs}F_{11}}{F_{1s}F_{r1}} = -\frac{\theta_2^{(11rs)}}{\theta_1^{(11rs)}},$$

can be uniquely recovered from  $\Psi^{(11rs)}$ .  $\square$

*Remark 7.3.* Note that we can relax [Assumptions 5](#) and [6](#) to the assumption of existence of  $i, j$  such that  $(\mathbf{G}_{1,:} \odot \mathbf{G}_{r,:} \odot \mathbf{H}_{1,:} \odot \mathbf{H}_{s,:}) \neq \mathbf{0}$  for all  $r \neq i$  and  $s \neq j$ .

### 7.3. Multilinear ranks of the core tensor.

*Proof of Proposition 4.11.* We will just prove the first statement as the second is analogous. The first unfolding of  $\mathcal{C}$  has the following block structure

$$\mathbf{C}^{(1)} = \begin{bmatrix} D_{\mathbf{G}}^{(1)} \mathbf{F} D_{\mathbf{H}}^{(1)} & D_{\mathbf{G}}^{(2)} \mathbf{F} D_{\mathbf{H}}^{(2)} & \cdots & D_{\mathbf{G}}^{(K)} \mathbf{F} D_{\mathbf{H}}^{(K)} \end{bmatrix}.$$

Let us select every  $j$ -th column of each block of  $\mathbf{C}^{(1)}$  and stack them into a submatrix  $\mathbf{Y} \in \mathbb{R}^{R \times K}$ . This submatrix has the form

$$\begin{aligned} \mathbf{Y} &= \begin{bmatrix} D_{\mathbf{G}}^{(1)} \mathbf{F}_{:,j} H_{j,1} & D_{\mathbf{G}}^{(2)} \mathbf{F}_{:,j} H_{j,2} & \cdots & D_{\mathbf{G}}^{(K)} \mathbf{F}_{:,j} H_{j,K} \end{bmatrix} \\ &= \text{Diag}\{\mathbf{F}_{:,j}\} [\mathbf{g}_1 H_{j,1} \quad \mathbf{g}_2 H_{j,2} \quad \cdots \quad \mathbf{g}_K H_{j,K}] = \text{Diag}\{\mathbf{F}_{:,j}\} (\mathbf{G} \odot \mathbf{H}_{j,:}) \end{aligned}$$

where  $\text{Diag}\{\mathbf{F}_{:,j}\}$  is nonsingular, hence  $\text{rank}\{\mathbf{Y}\} = R$  by assumption. This implies that  $\text{rank}\{\mathbf{C}^{(1)}\} = R$ .  $\square$

*Proof of Corollary 4.13.* 1. Let [Assumption 5](#) be satisfied. Then the matrix  $\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}$  has  $\binom{R+1}{2} \binom{S+1}{2}$  distinct rows, and, in particular, any collection of the distinct rows must be linearly independent. This implies that for any  $i, j$

$$(7.7) \quad \text{rank}\{\mathbf{G} \odot \mathbf{G}_{i,:} \odot \mathbf{H}_{j,:} \odot \mathbf{H}_{j,:}\} = R \text{ and } \text{rank}\{\mathbf{G}_{i,:} \odot \mathbf{G}_{i,:} \odot \mathbf{H}_{j,:} \odot \mathbf{H}\} = S.$$

Note that the matrices in (7.7) are just diagonally scaled versions of  $\mathbf{G} \odot \mathbf{H}_{j,:}$  and  $\mathbf{G}_{i,:} \odot \mathbf{H}$ , hence  $\text{rank}\{\mathbf{G} \odot \mathbf{H}_{j,:}\} = R$ ,  $\text{rank}\{\mathbf{G}_{i,:} \odot \mathbf{H}\} = S$  and thus the conditions of [Proposition 4.11](#) are satisfied.

2. If [Assumption 6](#) is satisfied, then the proof is very similar. In this case, the matrix  $\mathbf{G} \odot \mathbf{G} \odot \mathbf{G} \odot \mathbf{G}$  has  $\binom{R+3}{4}$  distinct rows, and, as in the previous case, we have that for any  $i$

$$(7.8) \quad \text{rank}\{\mathbf{G} \odot \mathbf{G}_{i,:} \odot \mathbf{G}_{i,:} \odot \mathbf{G}_{i,:}\} = R,$$

hence  $\text{rank}\{\mathbf{G} \odot \mathbf{G}_{i,:}\} = R$ , and the conditions [Proposition 4.11](#) are also satisfied (since  $\mathbf{G} = \mathbf{H}$ ).  $\square$

## 8. Proofs for the $2 \times 2$ case.

**8.1. Building intuition: the  $2 \times 2$  nonsymmetric case.** We provide a sketch proof of [Theorem 5.4](#) for a particular case  $R = S = 2$ . Assume that [Assumptions 1, 2, 5](#) and [7](#). Note that for  $R = 2, S = 2$  there is essentially one nontrivial matrix  $\Psi^{(ijrs)}$  that can be obtained, by choosing  $(i, r) = (j, s) = (1, 2)$ . It is easy to see that  $\Psi^{(1122)}$  is, in fact, a submatrix  $\Phi(\mathcal{C})$  then the condition [\(7.4\)](#) on  $\Psi^{(1122)}$  can be rewritten as

$$(\theta_1 \text{vec}\{e_1 \otimes e_1 \otimes e_2 \otimes e_2\} + \theta_2 \text{vec}\{e_1 \otimes e_2 \otimes e_2 \otimes e_1\})^\top \Phi(\mathcal{C}) = 0,$$

where  $\theta_1 = \theta_1^{(1122)}, \theta_2 = \theta_2^{(1122)}$  are the coefficients defined in [Lemma 7.2](#). Now let  $\tilde{\mathbf{a}}_k$  and  $\tilde{\mathbf{b}}_k$  be the rows of the inverses of  $\mathbf{A}$  and  $\mathbf{B}$  respectively<sup>7</sup>

$$\mathbf{A}^{-\top} = [\tilde{\mathbf{a}}_1 \quad \tilde{\mathbf{a}}_2], \quad \mathbf{B}^{-\top} = [\tilde{\mathbf{b}}_1 \quad \tilde{\mathbf{b}}_2].$$

Then by duality (i.e., using the fact that  $\Phi(\mathcal{T}) = (\mathbf{B} \boxtimes \mathbf{A} \boxtimes \mathbf{B} \boxtimes \mathbf{A})\Phi(\mathcal{C})$ ), we have that the vector

$$(8.1) \quad \mathbf{v} := (\theta_1 \text{vec}\{\tilde{\mathbf{a}}_1 \otimes \tilde{\mathbf{b}}_1 \otimes \tilde{\mathbf{a}}_2 \otimes \tilde{\mathbf{b}}_2\} + \theta_2 \text{vec}\{\tilde{\mathbf{a}}_1 \otimes \tilde{\mathbf{b}}_2 \otimes \tilde{\mathbf{a}}_2 \otimes \tilde{\mathbf{b}}_1\})$$

lies in the left kernel of  $\Phi(\mathcal{T})$  (i.e.  $\mathbf{v}^\top \Phi(\mathcal{T}) = 0$ ). Then the symmetrized vector is nonzero

$$\sigma_2(\mathbf{v}) = \theta_1 \sigma_2(\text{vec}\{\tilde{\mathbf{a}}_1 \otimes \tilde{\mathbf{a}}_1 \otimes \tilde{\mathbf{a}}_2 \otimes \tilde{\mathbf{a}}_2\}) + \theta_2 \sigma_2(\text{vec}\{\tilde{\mathbf{a}}_1 \otimes \tilde{\mathbf{a}}_2 \otimes \tilde{\mathbf{a}}_2 \otimes \tilde{\mathbf{a}}_1\}) \neq 0$$

and it lies in  $\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})$ . By [Lemma 5.3](#), this is, in fact, the only generator of the symmetrized kernel:

$$\sigma_2(\text{lker}\{\Phi(\mathcal{T})\}) = \text{Span}\{\sigma_2(\mathbf{v})\}.$$

Thus any vector  $\mathbf{p}$  generating  $\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})$  has the form  $c\sigma_2(\mathbf{v})$  for some nonzero constant. Now, applying the symmetrization and permutation map [\(5.2\)](#) leads to:

$$\mathbf{P} = \pi(c\sigma_2(\mathbf{v})) = c(\det \mathbf{F})\sigma_2(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2)(\sigma_2(\tilde{\mathbf{b}}_1 \boxtimes \tilde{\mathbf{b}}_2))^\top,$$

where we have used the fact that  $(\theta_1 + \theta_2) = \det \mathbf{F}$  by [Lemma 7.2](#) (which is nonzero by [Assumption 7](#)). This means that  $\mathbf{P}$  is a rank-1 matrix whose column and row spaces are

$$\text{range}\{\mathbf{P}\} = \text{Span}\{\sigma(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2)\}, \quad \text{range}\{\mathbf{P}^\top\} = \text{Span}\{\sigma(\tilde{\mathbf{b}}_1 \boxtimes \tilde{\mathbf{b}}_2)\}$$

Next, observe that

$$\langle \mathbf{a} \boxtimes \mathbf{a}, \sigma(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2) \rangle = \langle \mathbf{a}, \tilde{\mathbf{a}}_1 \rangle \langle \mathbf{a}, \tilde{\mathbf{a}}_2 \rangle,$$

which shows that both  $\mathbf{a}_1 \boxtimes \mathbf{a}_1$  and  $\mathbf{a}_2 \boxtimes \mathbf{a}_2$  are orthogonal to  $\sigma(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2)$ . These two vectors are linearly independent, and since  $\dim(\sigma_2(\text{lker}\{\mathbf{Q}\})) = 2$ , we get that

$$\sigma_2(\text{lker}\{\mathbf{P}\}) = \text{Span}\{\mathbf{a}_1 \boxtimes \mathbf{a}_1, \mathbf{a}_2 \boxtimes \mathbf{a}_2\}.$$

By similar manipulations, we can show that

$$\sigma_2(\text{lker}\{\mathbf{P}^\top\}) = \text{Span}\{\mathbf{b}_1 \boxtimes \mathbf{b}_1, \mathbf{b}_2 \boxtimes \mathbf{b}_2\},$$

which completes the proof of [Theorem 5.4](#) for a particular case  $R = S = 2$ .

Finally, we comment on how to simply obtain  $\mathbf{a}_1, \mathbf{a}_2$  from  $c\sigma(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2)$ . Assume for simplicity, that

$$\mathbf{a}_1 = [1 \quad x]^\top, \mathbf{a}_2 = [1 \quad y]^\top.$$

then  $\tilde{\mathbf{a}}_1 = c_1 [-x \quad 1]^\top, \tilde{\mathbf{a}}_2 = c_2 [-y \quad 1]^\top$  and thus

$$c\sigma(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2) = [z_0, z_1, z_1, z_2]^\top := c [xy \quad -(x+y)/2 \quad -(x+y)/2 \quad 1]^\top.$$

Now, given  $z_0, z_1, z_2$ , we can see that  $x, y$  must be the roots of the univariate polynomial  $z(t) = z_0 + 2z_1t + z_2t^2$ , thus the vectors  $\mathbf{a}_1, \mathbf{a}_2$  can be easily found from  $\sigma(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2)$ .

<sup>7</sup>These are the dual vectors,  $\langle \tilde{\mathbf{a}}_i, \mathbf{a}_j \rangle = \langle \tilde{\mathbf{b}}_i, \mathbf{b}_j \rangle = \delta_{ij}$

**8.2. Building intuition: the  $2 \times 2$  symmetric case.** Now let us consider the  $2 \times 2$  symmetric case, i.e.  $R = S = 2$ ,  $\mathbf{A} = \mathbf{B}$ ,  $\mathbf{G} = \mathbf{H}$ . Note that  $\mathbf{v}$  still lies in the left nullspace of  $\Phi(\mathcal{T})$ . Let us denote by  $\mathbf{w}$  the projection this vector on  $\text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^2)\})\}$  by double symmetrization, which has the form

$$\mathbf{w} = \theta_1 \sigma_2(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2) + \theta_2 \sigma_2(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2) \boxtimes \sigma_2(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2).$$

Then  $\text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\} \cap \text{lker}\{\Phi(\mathcal{T})\}$  is spanned by a single vector  $\mathbf{q} = c\mathbf{w}$ . Now take a projection of this vector on  $S^4(\mathbb{R}^2)$ , which is

$$\sigma_4(\mathbf{w}) = \sigma_4(\mathbf{v}) = (\det \mathbf{F}) \sigma_4(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2).$$

Let  $\mathbf{Q} \in \mathbb{R}^{2^3 \times 2}$  be the reshaping of  $\sigma_4(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2)$ , then the matrix  $\mathbf{Q}$  is of rank 2 because it has the following low-rank decomposition

$$\mathbf{Q} = \frac{1}{2} \left( \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2) \mathbf{a}_2^\top + \frac{1}{2} \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2) \mathbf{a}_1^\top \right).$$

Note that, on the one hand

$$\dim(\text{vec}\{S^3(\mathbb{R}^3)\} \cap \text{lker}\{\mathbf{Q}\}) = \dim(\text{vec}\{S^3(\mathbb{R}^3)\}) - \text{rank}\{\mathbf{Q}\} = 2,$$

and on the other hand, the two vectors  $\mathbf{a}_1 \boxtimes \mathbf{a}_1 \boxtimes \mathbf{a}_1$ ,  $\mathbf{a}_2 \boxtimes \mathbf{a}_2 \boxtimes \mathbf{a}_2$  are in the left nullspace of  $\mathbf{Q}$ . Therefore the symmetric left nullspace is spanned by those vectors.

*Remark 8.1.* The  $2 \times 2$  symmetric case has also interpretation in terms of univariate polynomials. Let  $\mathbf{a}_1, \mathbf{a}_2$  and  $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2$  be as in the previous subsection. Then the polynomial corresponding to the vector  $\sigma_4(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2)$

$$z(t) = \left\langle \sigma_4(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2), \begin{bmatrix} 1 \\ t \end{bmatrix}^{\boxtimes 4} \right\rangle = \left( \left\langle \tilde{\mathbf{a}}_1, \begin{bmatrix} 1 \\ t \end{bmatrix} \right\rangle \right)^2 \left( \left\langle \tilde{\mathbf{a}}_2, \begin{bmatrix} 1 \\ t \end{bmatrix} \right\rangle \right)^2 = c_1^2 c_2^2 (t-x)^2 (t-y)^2$$

thus it has two double roots. Therefore  $x, y$  can be found just by finding the roots of the polynomial  $\sqrt{z(t)}$ .

## 9. Nonsymmetric case: proof of the main theorem.

**9.1. Basis of the left kernel.** Similarly to Subsection 8.1, we can get the basis of the symmetric kernel of the structured matrix  $\Phi(\mathcal{T})$  in the general case  $R, S \geq 2$ . Let

$$(9.1) \quad \tilde{\mathbf{A}} = \mathbf{A}^{-\top} = [\tilde{\mathbf{a}}_1 \quad \cdots \quad \tilde{\mathbf{a}}_R], \quad \tilde{\mathbf{B}} = \mathbf{B}^{-\top} = [\tilde{\mathbf{b}}_1 \quad \cdots \quad \tilde{\mathbf{b}}_S],$$

so that  $\{\tilde{\mathbf{a}}_k\}_{k=1}^R, \{\tilde{\mathbf{b}}_k\}_{k=1}^S$  are the dual vectors for  $\{\mathbf{a}_k\}_{k=1}^R$  and  $\{\mathbf{b}_k\}_{k=1}^S$ , respectively. Then the following lemma generalized the derivations in Subsection 8.1.

**PROPOSITION 9.1.** *Let  $\mathcal{T}$  have a PT2D with invertible  $\mathbf{A}, \mathbf{B}$ . For each pair  $1 \leq i, r \leq R, 1 \leq j, s \leq S$ , define the following vector  $\mathbf{v}^{(ijrs)} \in \sigma(\mathbb{R}^{SRSR})$ :*

$$(9.2) \quad \mathbf{v}^{(ijrs)} := \left( \theta_1^{(ijrs)} \text{vec}\{\tilde{\mathbf{a}}_i \otimes \tilde{\mathbf{b}}_j \otimes \tilde{\mathbf{a}}_r \otimes \tilde{\mathbf{b}}_s\} + \theta_2^{(ijrs)} \text{vec}\{\tilde{\mathbf{a}}_i \otimes \tilde{\mathbf{b}}_s \otimes \tilde{\mathbf{a}}_r \otimes \tilde{\mathbf{b}}_j\} \right),$$

where  $\theta_1^{(ijrs)}, \theta_2^{(ijrs)}$  are defined in (7.3). Then we have that:

1. Each vector  $\mathbf{v}^{(ijrs)}$  belongs to the left nullspace of  $\Phi(\mathcal{T})$ ;
2. Under Assumptions 1, 2 and 5, the symmetrizations of  $\binom{R}{2} \binom{S}{2}$  vectors corresponding to distinct pairs of  $(i, r)$  and  $(j, s)$  form a basis of the symmetric left kernel of  $\Phi(\mathcal{T})$ , i.e.,

$$\sigma_2(\text{lker}\{\Phi(\mathcal{T})\}) = \text{Span}\left\{ \left\{ \sigma_2(\mathbf{v}^{(ijrs)}) \right\}_{\substack{1 \leq i < r \leq R \\ 1 \leq j < s \leq S}} \right\}$$

3. The permutation/symmetrization map (5.2) applied to  $\mathbf{v}^{(ijrs)}$  gives the  $R^2 \times S^2$  matrix of rank at most 1

$$\pi(\sigma_2(\mathbf{v}^{(ijrs)})) = \pi(\mathbf{v}^{(ijrs)}) = (\det \mathbf{F}_{(i,r),(j,s)}) \cdot \sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r) (\sigma_2(\tilde{\mathbf{b}}_j \boxtimes \tilde{\mathbf{b}}_s))^\top$$

*Proof.* 1. The proof repeats the derivations in [Subsection 8.1](#), by substituting  $\tilde{\mathbf{a}}_i, \tilde{\mathbf{a}}_r, \tilde{\mathbf{b}}_j, \tilde{\mathbf{b}}_s$  instead of  $\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_1, \tilde{\mathbf{b}}_1, \tilde{\mathbf{b}}_2$  in (8.1). In particular we use the fact that

$$(9.3) \quad \mathbf{v}^{(ijrs)} = (\tilde{\mathbf{B}} \boxtimes \tilde{\mathbf{A}} \boxtimes \tilde{\mathbf{B}} \boxtimes \tilde{\mathbf{A}}) \left( \theta_1^{(ijrs)} \text{vec}\{\mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_r \otimes \mathbf{e}_s\} + \theta_2^{(ijrs)} \text{vec}\{\mathbf{e}_i \otimes \mathbf{e}_s \otimes \mathbf{e}_r \otimes \mathbf{e}_j\} \right).$$

2. Thanks to (9.3), we have that  $\sigma_2(\mathbf{v}^{(ijrs)}) = (\tilde{\mathbf{B}} \boxtimes \tilde{\mathbf{A}} \boxtimes \tilde{\mathbf{B}} \boxtimes \tilde{\mathbf{A}}) \mathbf{w}^{(ijrs)}$ , where

$$\begin{aligned} \mathbf{w}^{(ijrs)} &= \frac{\theta_1^{(ijrs)}}{2} (\text{vec}\{\mathbf{e}_i \otimes \mathbf{e}_j \otimes \mathbf{e}_r \otimes \mathbf{e}_s\} + \text{vec}\{\mathbf{e}_r \otimes \mathbf{e}_s \otimes \mathbf{e}_i \otimes \mathbf{e}_j\}) + \\ &\quad \frac{\theta_2^{(ijrs)}}{2} (\text{vec}\{\mathbf{e}_i \otimes \mathbf{e}_s \otimes \mathbf{e}_r \otimes \mathbf{e}_j\} + \text{vec}\{\mathbf{e}_r \otimes \mathbf{e}_j \otimes \mathbf{e}_i \otimes \mathbf{e}_s\}). \end{aligned}$$

We note that we take only unique pairs of indices  $1 \leq i < r \leq R$ ,  $1 \leq j < s \leq S$ , hence for another two pairs  $1 \leq i' < r' \leq R$ ,  $1 \leq j' < s' \leq S$ , such that  $(i, r) \neq (i', r')$  or  $(j, s) \neq (j', s')$ , we have  $\langle \mathbf{w}^{(ijrs)}, \mathbf{w}^{(i'j'r's')} \rangle = 0$ . Therefore, the  $\mathbf{w}^{(ijrs)}$  are mutually ortogonal and the vectors  $\{\mathbf{v}^{(ijrs)}\}_{\substack{1 \leq i < r \leq R \\ 1 \leq j < s \leq S}}$  are linearly independent.

3. Consider  $\pi$  applied to  $\mathbf{w}^{(ijrs)}$ , which then becomes

$$\begin{aligned} \pi(\mathbf{w}^{(ijrs)}) &= \frac{\theta_1^{(ijrs)} + \theta_2^{(ijrs)}}{4} \left( (\mathbf{e}_i \boxtimes \mathbf{e}_r)(\mathbf{e}_j \boxtimes \mathbf{e}_s)^\top + (\mathbf{e}_r \boxtimes \mathbf{e}_i)(\mathbf{e}_s \boxtimes \mathbf{e}_j)^\top + \right. \\ &\quad \left. (\mathbf{e}_i \boxtimes \mathbf{e}_r)(\mathbf{e}_s \boxtimes \mathbf{e}_j)^\top + (\mathbf{e}_r \boxtimes \mathbf{e}_i)(\mathbf{e}_j \boxtimes \mathbf{e}_s)^\top \right) \\ &= \sigma_2(\mathbf{e}_i \boxtimes \mathbf{e}_r)(\sigma_2(\mathbf{e}_j \boxtimes \mathbf{e}_s))^\top. \end{aligned}$$

By multilinearity we have

$$\pi((\tilde{\mathbf{B}} \boxtimes \tilde{\mathbf{A}} \boxtimes \tilde{\mathbf{B}} \boxtimes \tilde{\mathbf{A}}) \mathbf{w}^{(ijrs)}) = (\tilde{\mathbf{A}} \boxtimes \tilde{\mathbf{A}}) \pi(\mathbf{w}^{(ijrs)}) (\tilde{\mathbf{B}} \boxtimes \tilde{\mathbf{B}})^\top,$$

which completes the proof.  $\square$

**9.2. Extracting  $A$  and  $B$  from the kernel.** We begin by proving the following lemma.

LEMMA 9.2. For  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H}) \in \mathbb{R}^{R \times S \times K}$  with  $\mathbf{A} \in \mathbb{R}^{R \times R}$  and  $\mathbf{B} \in \mathbb{R}^{S \times S}$  we have:

$$(9.4) \quad \text{rank}\{\Phi(\mathcal{T})\} \leq \text{rank}\{\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}\}.$$

*Proof.* By the properties of vectorizations,  $\text{vec}\{\mathcal{T}_{::,k}\} = (\mathbf{B} \boxtimes \mathbf{A}) \text{Diag}\{\text{vec}\{\mathbf{F}\}\}(\mathbf{h}_k \boxtimes \mathbf{g}_k)$ , hence the columns  $(\Phi(\mathcal{T}))_{:,k} = \text{vec}\{\mathcal{T}_{::,k}\} \boxtimes \text{vec}\{\mathcal{T}_{::,k}\}$  can be expressed as

$$(\Phi(\mathcal{T}))_{:,k} = (\mathbf{B} \boxtimes \mathbf{A} \boxtimes \mathbf{B} \boxtimes \mathbf{A}) \text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\}(\mathbf{h}_k \boxtimes \mathbf{g}_k \boxtimes \mathbf{h}_k \boxtimes \mathbf{g}_k),$$

which is equivalent to saying that

$$(9.5) \quad \Phi(\mathcal{T}) = (\mathbf{B} \boxtimes \mathbf{A} \boxtimes \mathbf{B} \boxtimes \mathbf{A}) \text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\}(\mathbf{H} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{G}).$$

Then the statement follows since  $(\mathbf{B} \boxtimes \mathbf{A} \boxtimes \mathbf{B} \boxtimes \mathbf{A}) \text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\}$  is a square matrix and

$$\text{rank}\{\mathbf{H} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{G}\} = \text{rank}\{\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}\}. \quad \square$$

Lemma [Lemma 9.2](#) serves as a base for proving [Lemma 5.3](#).

*Proof of Lemma 5.3.* 1. The columns of  $\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}$  are vectorizations of tensors  $(\mathbf{g}_k \otimes \mathbf{g}_k) \otimes (\mathbf{h}_k \otimes \mathbf{h}_k)$ ,  
hence

$$\text{range}\{\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}\} \subseteq \text{vec}\{S^2(\mathbb{R}^R) \otimes S^2(\mathbb{R}^S)\}.$$

This implies by [Lemma 9.2](#),

$$(9.6) \quad \text{rank}\{\Phi(\mathcal{T})\} \leq \text{rank}\{\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}\} \leq \dim\{S^2(\mathbb{R}^R) \otimes S^2(\mathbb{R}^S)\} = \binom{R+1}{2} \binom{S+1}{2},$$



2. Let [Assumptions 1](#) and [2](#) be satisfied. Then, by inspecting the proof of [Lemma 9.2](#), we see that in (9.5),  $(\mathbf{B} \boxtimes \mathbf{A} \boxtimes \mathbf{B} \boxtimes \mathbf{A}) \text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\}$  is an invertible matrix and hence equality  $\text{rank}\{\Phi(\mathcal{T})\} = \text{rank}\{\mathbf{G} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{H}\}$  is achieved. Therefore (9.6) becomes an equality if, in addition, [Assumption 5](#) holds true.

The dimension of the symmetrized kernel follows from a simple count of dimensions, as

$$\begin{aligned} \dim(\sigma_2(\text{lker}\{\Phi(\mathcal{T})\})) &= \dim(S^2(\mathbb{R}^{RS})) - \text{rank}\{\Phi(\mathcal{T})\} \\ &= \binom{RS+1}{2} - \binom{R+1}{2} \binom{S+1}{2} = \binom{R}{2} \binom{S}{2}. \end{aligned}$$

3. Assume, that [Assumptions 2](#) and [5](#) are not satisfied simultaneously. We have the following cases:
- If [Assumption 5](#) is not satisfied, then the strict inequality in (9.6) follows.
  - If [Assumption 2](#) is not satisfied, there is  $i, j$  such that  $F_{i,j} = 0$ , hence the (only) diagonal element in  $\text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\}$  corresponding to  $F_{i,j}^2$  is zero (this diagonal element occurs at the position of 1 in the vector  $\mathbf{v} = \mathbf{e}_j \boxtimes \mathbf{e}_i \boxtimes \mathbf{e}_j \boxtimes \mathbf{e}_i$ ). Denote  $\Pi = \mathbf{I}_{(RS)^2} - \mathbf{v}\mathbf{v}^\top$  (the identity matrix without the corresponding diagonal element). Then

$$\begin{aligned} \text{rank}\{\Phi(\mathcal{T})\} &\leq \text{rank}\{\text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\}(\mathbf{H} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{G})\} \\ &\leq \text{rank}\{\Pi(\mathbf{H} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{G})\} < \text{rank}\{(\mathbf{H} \odot \mathbf{G} \odot \mathbf{H} \odot \mathbf{G})\}, \end{aligned}$$

where the strict inequality follows from the fact that  $\Pi((\mathbf{h} \boxtimes \mathbf{g} \boxtimes \mathbf{h} \boxtimes \mathbf{g}))$  does not contain the element  $h_i^2 g_i^2$  which appears only once in  $\mathbf{h} \boxtimes \mathbf{g} \boxtimes \mathbf{h} \boxtimes \mathbf{g}$ , for all  $\mathbf{g} \in \mathbb{R}^R$ ,  $\mathbf{h} \in \mathbb{R}^S$ .  $\square$

*Proof of Theorem 5.4.*

1. Let  $\mathbf{P}_1, \dots, \mathbf{P}_M$  be as in [Theorem 5.4](#). Then we have that, by [Proposition 9.1](#)

$$\text{Span}\{\{\mathbf{P}_k\}_{k=1}^M\} = \text{Span}\left\{\left\{(\det \mathbf{F}_{(i,r),(j,s)}) \cdot \sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r)(\sigma_2(\tilde{\mathbf{b}}_j \boxtimes \tilde{\mathbf{b}}_s))^\top\right\}_{\substack{1 \leq i \leq r \leq R \\ 1 \leq j \leq s \leq S}}\right\}.$$

Therefore, we have that

$$(9.7) \quad \text{range}\{\mathbf{P}_A\} = \text{Span}\{\{\text{range}\{\mathbf{P}_k\}\}_{k=1}^M\} \subseteq \text{Span}\{\{\sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r)\}_{1 \leq i \leq r \leq R}\},$$

and, by the dimension count,

$$(9.8) \quad \dim(\sigma_2(\text{lker}\{\mathbf{P}_A\})) = \dim S^2(\mathbb{R}^R) - \text{rank}\{\mathbf{P}_A\} \geq \binom{R+1}{2} - \binom{R}{2} = R.$$

If  $\text{krank}\{\mathbf{F}^\top\} \geq 2$ , for any pair  $(i, r)$  there exist  $(j, s)$  such that  $\det \mathbf{F}_{(i,r),(j,s)} \neq 0$ , then this implies that the equality holds in (9.7) and (9.8). On the other hand, if  $\text{krank}\{\mathbf{F}^\top\} = 1$ , then there exists a pair  $(i, j)$ , such that  $\sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r) \notin \text{range}\{\mathbf{P}_A\}$ , hence  $\dim(\sigma_2(\text{lker}\{\mathbf{P}_A\})) > R$  in this case. Finally we note that for any  $i, r, \ell$

$$\langle \sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r), \mathbf{a}_\ell \boxtimes \mathbf{a}_\ell \rangle = \langle \tilde{\mathbf{a}}_i, \mathbf{a}_\ell \rangle \langle \tilde{\mathbf{a}}_r, \mathbf{a}_\ell \rangle = 0.$$

Therefore,

$$\text{Span}\{\{\mathbf{a}_\ell \boxtimes \mathbf{a}_\ell\}_{\ell=1}^R\} \subseteq \sigma_2(\text{lker}\{\mathbf{P}_A\}),$$

where the equality holds if and only if  $\text{krank}\{\mathbf{F}^\top\} \geq 2$ .

2. The proof is analogous to the one of the previous statement. We observe that

$$(9.9) \quad \text{range}\{\mathbf{P}_B\} \subseteq \text{Span}\{\{\sigma_2(\tilde{\mathbf{b}}_j \boxtimes \tilde{\mathbf{b}}_s)\}_{1 \leq j \leq s \leq S}\}.$$

and

$$\text{Span}\{\{\mathbf{b}_\ell \boxtimes \mathbf{b}_\ell\}_{\ell=1}^R\} \subseteq \sigma_2(\text{lker}\{\mathbf{P}_B\}),$$

where the equalities hold if and only if  $\text{krank}\{\mathbf{F}\} = 2$ , by invoking the dimension count argument.  $\square$

**9.3. On the case of collinear columns or rows of  $\mathbf{F}$ .** We will make some comments on the case of collinear rows or columns in  $\mathbf{F}$  (see [Subsection 4.5](#)). In fact, if [Assumption 7](#) is not satisfied, then we will show that  $\mathbf{A}$  and  $\mathbf{B}$  can be recovered subject to additional ambiguities outlined in [Subsection 4.5](#). To do this, we use the following proposition

**PROPOSITION 9.3.** *Let [Assumptions 1, 2 and 5](#) be satisfied,  $\mathbf{F}$  be as in (4.7) (see [Lemma 4.15](#)). Then for  $\mathbf{P}_A$  and  $\mathbf{P}_B$  as in [Theorem 5.4](#), it holds that*

$$(9.10) \quad \dim(\sigma_2(\text{lker}\{\mathbf{P}_A\})) = \text{Span}\{(\{\mathbf{A} \text{Diag}\{\mathbf{Y}_1, \dots, \mathbf{Y}_{R'}\} \mathbf{A}^\top : \mathbf{Y}_k \text{ is } m_k \times m_k \text{ symmetric}\})\},$$

$$(9.11) \quad \dim(\sigma_2(\text{lker}\{\mathbf{P}_B\})) = \text{Span}\{(\{\mathbf{B} \text{Diag}\{\mathbf{Z}_1, \dots, \mathbf{Z}_{S'}\} \mathbf{B}^\top : \mathbf{Z}_k \text{ is } n_k \times n_k \text{ symmetric}\})\}.$$

*Proof.* We will prove the statement just for  $\mathbf{A}$ . We define  $\mathcal{J}_k$ ,  $1 \leq k \leq R'$ , such that

$$\begin{aligned} \mathcal{J}_1 &= \{1, \dots, m_1\} \\ \mathcal{J}_2 &= \{m_1 + 1, \dots, m_1 + m_2\} \\ &\vdots \\ \mathcal{J}_{R'} &= \left\{ \sum_{\ell=1}^{R'} m_\ell + 1, \dots, R \right\}. \end{aligned}$$

The sets of indices  $\mathcal{J}_k$  correspond to blocks of rows in (4.7). Then, as in the proof of [Theorem 5.4](#),  $\sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r)$  belongs to  $\text{range}\{\mathbf{P}_A\}$  if and only if  $\text{rank}\{\mathbf{F}_{(i,r),:}\} = 2$ , which can only happen in  $i$  and  $r$  belong to different blocks of rows. Therefore, we have that

$$(9.12) \quad \text{range}\{\mathbf{P}_A\} = \text{Span}\{ \{ \sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r) \}_{i \in \mathcal{J}_k, r \in \mathcal{J}_\ell, k \neq \ell} \}.$$

Note that the dimension of this space is equal to

$$\dim(\text{range}\{\mathbf{P}_A\}) = \binom{R+1}{2} - \left( \sum_{\ell=1}^{S'} \binom{m_\ell + 1}{2} \right).$$

as we only forbid the pairs belonging to the same blocks. Next, we observe that for any  $(i', r')$  and  $(i, r)$

$$(9.13) \quad \langle \sigma_2(\mathbf{a}_{i'} \boxtimes \mathbf{a}_{r'}), \sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r) \rangle = \frac{1}{2} (\langle \mathbf{a}_{i'}, \tilde{\mathbf{a}}_i \rangle \langle \mathbf{a}_{r'}, \tilde{\mathbf{a}}_r \rangle + \langle \mathbf{a}_{i'}, \tilde{\mathbf{a}}_r \rangle \langle \mathbf{a}_{r'}, \tilde{\mathbf{a}}_i \rangle).$$

Now, if  $i', r' \in \mathcal{J}_t$  and  $i \in \mathcal{J}_k, r \in \mathcal{J}_\ell, k \neq \ell$ , then we have that (9.13) vanishes because  $(i', r') \notin \{(i, r), (r, i)\}$  due to the constraints. This implies that for a fixed  $1 \leq t \leq R$ , vectorization of any matrix

$$\mathbf{X} = \sum_{i, r \in \mathcal{J}_t} c_{i,j} \mathbf{a}_i \mathbf{a}_r^\top.$$

belongs to the left kernel:  $\text{vec}\{\mathbf{X}\} \in \text{lker}\{\mathbf{P}_A\}$ . Thus any matrix of the form as in (9.10) belongs to  $\sigma_2(\text{lker}\{\mathbf{P}_A\})$  and the two sets in (9.10) coincide by the dimension count.  $\square$

## 10. Symmetric case: proof of the main theorem.

**10.1. Basis of the left kernel.** Let  $\mathbf{a}_k$  and  $\tilde{\mathbf{a}}_k$  be in [Section 9](#). In the symmetric case we have  $\mathbf{a}_k = \mathbf{b}_k$  and  $\tilde{\mathbf{a}}_k = \tilde{\mathbf{b}}_k$ . Note that the vector  $\mathbf{v}^{(ijrs)}$  belongs to the left kernel of  $\Phi(\mathcal{T})$  and in this case becomes

$$(10.1) \quad \mathbf{v}^{(ijrs)} := \left( \theta_1^{(ijrs)} \text{vec}\{\tilde{\mathbf{a}}_i \otimes \tilde{\mathbf{a}}_j \otimes \tilde{\mathbf{a}}_r \otimes \tilde{\mathbf{a}}_s\} + \theta_2^{(ijrs)} \text{vec}\{\tilde{\mathbf{a}}_i \otimes \tilde{\mathbf{a}}_s \otimes \tilde{\mathbf{a}}_r \otimes \tilde{\mathbf{a}}_j\} \right).$$

Let  $\sigma_{(2,2)}(\cdot)$  be the orthogonal projection on  $\text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^S)\})\}$ . Then we have that

$$\mathbf{u}^{(ijrs)} := \sigma_{(2,2)}(\mathbf{v}^{(ijrs)}) = \theta_1^{(ijrs)} \sigma_2(\sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_j) \boxtimes \sigma_2(\tilde{\mathbf{a}}_r \boxtimes \tilde{\mathbf{a}}_s)) + \theta_2^{(ijrs)} \sigma_2(\sigma_2(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_s) \boxtimes \sigma_2(\tilde{\mathbf{a}}_r \boxtimes \tilde{\mathbf{a}}_j)).$$

We have the following proposition.

PROPOSITION 10.1. Let  $\mathcal{T}$  have a symmetric PT2D with invertible  $\mathbf{A} = \mathbf{B}$ . Then we have that:

1. Under [Assumptions 1, 2 and 6](#), the symmetric left kernel is spanned by the vectors  $\mathbf{u}^{(ijrs)}$

$$\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\}) = \text{Span}\{\{\mathbf{u}^{(ijrs)}\}_{\substack{1 \leq i < r \leq R \\ 1 \leq j < s \leq R}}\}$$

2. The full symmetrization applied to  $\mathbf{u}^{(ijrs)}$  gives the following

$$\sigma_4(\mathbf{u}^{(ijrs)}) = \sigma_4(\mathbf{v}^{(ijrs)}) = (\det \mathbf{F}_{(i,r),(j,s)}) \cdot \sigma_4(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r \boxtimes \tilde{\mathbf{a}}_j \boxtimes \tilde{\mathbf{a}}_s).$$

*Proof.* 1. Note that we have

$$\text{Span}\{\{\mathbf{u}^{(ijrs)}\}_{\substack{1 \leq i < r \leq R \\ 1 \leq j < s \leq R}}\} \subseteq \sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\}).$$

To show that the inequality holds, we will show that we can choose  $N = \frac{R^2(R-1)(R+1)}{12}$  linearly independent vectors among  $\mathbf{u}^{(ijrs)}$ . Consider the following sets of indices:

$$(10.2) \quad \mathcal{J}_2 = \{(k, k, \ell, \ell)\}_{1 \leq k < \ell \leq R},$$

$$(10.3) \quad \mathcal{J}_3 = \{(k, \ell, m, m), (k, \ell, \ell, m), (k, k, \ell, m)\}_{1 \leq k < \ell < m \leq R},$$

$$(10.4) \quad \mathcal{J}_4 = \{(k, \ell, m, n), (k, m, \ell, n)\}_{1 \leq k < \ell < m < n \leq R},$$

Next, as in the proof [Proposition 9.1](#), we express  $\mathbf{u}^{(ijrs)} = (\tilde{\mathbf{A}} \boxtimes \tilde{\mathbf{A}} \boxtimes \tilde{\mathbf{A}} \boxtimes \tilde{\mathbf{A}})\mathbf{x}^{(ijrs)}$ , where

$$\begin{aligned} \mathbf{x}^{(ijrs)} &= \theta_1^{(ijrs)} \sigma_2(\sigma_2(\mathbf{e}_i \boxtimes \mathbf{e}_j) \boxtimes \sigma_2(\mathbf{e}_r \boxtimes \mathbf{e}_s)) + \theta_2^{(ijrs)} \sigma_2(\sigma_2(\mathbf{e}_i \boxtimes \mathbf{e}_s) \boxtimes \sigma_2(\mathbf{e}_r \boxtimes \mathbf{e}_j)) \\ &= \frac{\theta_1^{(ijrs)}}{4} ((\mathbf{e}_i \boxtimes \mathbf{e}_j + \mathbf{e}_j \boxtimes \mathbf{e}_i) \boxtimes (\mathbf{e}_r \boxtimes \mathbf{e}_s + \mathbf{e}_s \boxtimes \mathbf{e}_r) + (\mathbf{e}_r \boxtimes \mathbf{e}_s + \mathbf{e}_s \boxtimes \mathbf{e}_r) \boxtimes (\mathbf{e}_i \boxtimes \mathbf{e}_j + \mathbf{e}_j \boxtimes \mathbf{e}_i)) \\ &= \frac{\theta_2^{(ijrs)}}{4} ((\mathbf{e}_i \boxtimes \mathbf{e}_s + \mathbf{e}_s \boxtimes \mathbf{e}_i) \boxtimes (\mathbf{e}_r \boxtimes \mathbf{e}_j + \mathbf{e}_j \boxtimes \mathbf{e}_r) + (\mathbf{e}_r \boxtimes \mathbf{e}_j + \mathbf{e}_j \boxtimes \mathbf{e}_r) \boxtimes (\mathbf{e}_i \boxtimes \mathbf{e}_s + \mathbf{e}_s \boxtimes \mathbf{e}_i)). \end{aligned}$$

We note that all the multiindices in  $\mathcal{J}_2, \mathcal{J}_3, \mathcal{J}_4$  correspond to different choices of 4 numbers (with possible repetition) among  $\{1, \dots, R\}$  (except the set  $\mathcal{J}_4$  which is composed of couple of multi-indices, corresponding to the same subset of numbers). Therefore, for two distinct multi-indices  $(i, j, r, s), (i', j', r', s') \in \mathcal{J}_2 \cup \mathcal{J}_3 \cup \mathcal{J}_4$ , we have  $\langle \mathbf{x}^{(ijrs)}, \mathbf{x}^{(i'j'r's')} \rangle = 0$  if

$$(i, j, r, s) \in \mathcal{J}_t, (i', j', r', s') \in \mathcal{J}_{t'}, \text{ with } (t, t') \in \{(2, 2), (2, 3), (2, 4), (3, 3), (3, 4)\}.$$

This proves that the vectors in  $\{\mathbf{x}^{(ijrs)}\}_{(i,j,r,s) \in \mathcal{J}_2 \cup \mathcal{J}_3}$  are mutually orthogonal and that

$$\{\mathbf{x}^{(ijrs)}\}_{(i,j,r,s) \in \mathcal{J}_2 \cup \mathcal{J}_3} \perp \{\mathbf{x}^{(ijrs)}\}_{(i,j,r,s) \in \mathcal{J}_4}.$$

Next, we are going to show that  $\{\mathbf{x}^{(ijrs)}\}_{(i,j,r,s) \in \mathcal{J}_4}$  is linearly independent. For this we note that for two multi-indices  $(i, j, r, s), (i', j', r', s') \in \mathcal{J}_4$  we have  $\langle \mathbf{x}^{(ijrs)}, \mathbf{x}^{(i'j'r's')} \rangle = 0$  as long as they are not linked by a permutation (i.e.,  $\{i, j, r, s\} \neq \{i', j', r', s'\}$ ). Therefore, it is left to show that  $\mathbf{x}^{(klmn)}$  is linearly independent of  $\mathbf{x}^{(kmln)}$ . For simplicity, take  $(k, \ell, m, n) = (1, 2, 3, 4)$  and note that the term  $\mathbf{e}_1 \boxtimes \mathbf{e}_2 \boxtimes \mathbf{e}_3 \boxtimes \mathbf{e}_4$  appears only in the expansion of  $\mathbf{x}^{(1234)}$  while  $\mathbf{e}_1 \boxtimes \mathbf{e}_3 \boxtimes \mathbf{e}_2 \boxtimes \mathbf{e}_4$  appears only in the expansion of  $\mathbf{x}^{(1324)}$ . This proves that  $\mathbf{x}^{(1234)}$  and  $\mathbf{x}^{(1324)}$  are linearly independent.

Finally, we note that the cardinality of the set

$$\#\{\mathcal{J}_2 \cup \mathcal{J}_2 \cup \mathcal{J}_3\} = \binom{R}{2} + 3\binom{R}{3} + 2\binom{R}{4} = \frac{R^2(R-1)(R+1)}{12} = \dim(\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\})).$$

where the last equality holds from [Lemma 6.3](#). This proves that  $\{\mathbf{u}^{(ijrs)}\}_{(i,j,r,s) \in \mathcal{J}_2 \cup \mathcal{J}_3 \cup \mathcal{J}_4}$  forms the basis of  $\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\})$ .

2. The proof is similar to the corresponding statement of [Proposition 9.1](#). Note that  $\text{vec}\{S^4(\mathbb{R}^R)\} \subset \text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\}$  and therefore

$$\sigma_4(\mathbf{u}^{(ijrs)}) = \sigma_4(\sigma_{(2,2)}(\mathbf{v}^{(ijrs)})) = \sigma_4(\mathbf{v}^{(ijrs)}).$$

Then it is easy to see that

$$\begin{aligned} \sigma_4(\mathbf{v}^{(ijrs)}) &= \theta_1^{(ijrs)} \sigma_4(\text{vec}\{\tilde{\mathbf{a}}_i \otimes \tilde{\mathbf{a}}_j \otimes \tilde{\mathbf{a}}_r \otimes \tilde{\mathbf{a}}_s\}) + \theta_2^{(ijrs)} \sigma_4(\text{vec}\{\tilde{\mathbf{a}}_i \otimes \tilde{\mathbf{a}}_s \otimes \tilde{\mathbf{a}}_r \otimes \tilde{\mathbf{a}}_j\}) \\ &= (\theta_1^{(ijrs)} + \theta_2^{(ijrs)}) \sigma_4(\text{vec}\{\tilde{\mathbf{a}}_i \otimes \tilde{\mathbf{a}}_j \otimes \tilde{\mathbf{a}}_r \otimes \tilde{\mathbf{a}}_s\}) = (\det \mathbf{F}_{(i,r),(j,s)}) \sigma_4(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_j \boxtimes \tilde{\mathbf{a}}_r \boxtimes \tilde{\mathbf{a}}_s) \square \end{aligned}$$

### 10.2. Extracting the vectors from the left kernel.

*Proof of Lemma 6.3.* 1. Similarly, the columns of  $\mathbf{G}^{\odot 4} = \mathbf{G} \odot \mathbf{G} \odot \mathbf{G} \odot \mathbf{G}$  are vectorizations of symmetric tensors ( $\mathbf{g}^{\otimes 4}$ ), hence by Lemma 9.2 we have

$$(10.5) \quad \text{rank}\{\Phi(\mathcal{T})\} \leq \text{rank}\{\mathbf{G}^{\odot 4}\} \leq \dim\{S^4(\mathbb{R}^R)\} = \binom{R+3}{4},$$

The second part of the statement follows from the dimension count:

$$\begin{aligned} \dim(\sigma_{(2,2)}(\text{lker}\{\Phi(\mathcal{T})\})) &= \dim(\text{vec}\{S^2(\text{vec}\{S^2(\mathbb{R}^R)\})\}) - \text{rank}\{\Phi(\mathcal{T})\} \\ &= \frac{((\binom{R+1}{2} + 1)\binom{R+1}{2})}{2} - \binom{R+3}{4} = \frac{R^2(R-1)(R+1)}{12}. \end{aligned}$$

2. As in the nonsymmetric case, both inequalities in (10.5) become equalities if all Assumptions 1, 2 and 6 are satisfied.
3. We proceed as in the nonsymmetric case. Assume that Assumptions 2 and 6 are not satisfied simultaneously. We have the following cases:

- If Assumption 6 is not satisfied, then the strict inequality in (10.5) follows.
- If Assumption 2 is not satisfied, there is  $i, j$  such that  $F_{j,i} = F_{i,j} = 0$ . Then we have that  $F_{j,i}F_{i,i} = F_{i,j}F_{i,i} = 0$  and all the corresponding diagonal elements in  $\text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\}$  (there are 4 of them if  $i \neq j$  and just one otherwise) are zero. This implies that in the product  $\text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\} \mathbf{g}^{\boxtimes 4}$  the monomial  $g_i^3 g_j$  never appears and thus the corresponding rows in

$$\text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\} \mathbf{G}^{\odot 4}$$

are zero. Therefore,

$$\text{rank}\{\Phi(\mathcal{T})\} \leq \text{rank}\{\text{Diag}\{\text{vec}\{\mathbf{F}\} \boxtimes \text{vec}\{\mathbf{F}\}\}(\mathbf{G}^{\odot 4})\} < \text{rank}\{(\mathbf{G}^{\odot 4})\},$$

which completes the proof.  $\square$

*Proof of Theorem 6.4.* Let  $\mathbf{p}_1, \dots, \mathbf{p}_N$  be as in Theorem 6.4. Then we have that, by Proposition 10.1

$$\text{Span}\{\{\mathbf{p}_k\}_{k=1}^N\} = \text{Span}\left\{\left\{(\det \mathbf{F}_{(i,r),(j,s)}) \cdot \sigma_4(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r \boxtimes \tilde{\mathbf{a}}_j \boxtimes \tilde{\mathbf{a}}_s)\right\}^{\top}_{\substack{1 \leq i < r \leq R \\ 1 \leq j < s \leq R}}\right\}.$$

Now, let  $\mathbf{P}_k = \text{matr}_{R^3, R}\{\mathbf{p}_k\}$  and  $\mathbf{P}_{sym}$  as in (6.2). Then we have that

$$(10.6) \quad \text{range}\{\mathbf{P}_{sym}\} = \text{Span}\left\{\left\{\text{range}\{\mathbf{Z}^{(ijrs)}\}\right\}_{\substack{(i,j,r,s): \\ \det \mathbf{F}_{(i,r),(j,s)} \neq 0}}\right\}$$

where  $\mathbf{Z}^{(ijrs)} = \text{matr}_{R^3, R}\{\sigma_4(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_j \boxtimes \tilde{\mathbf{a}}_r \boxtimes \tilde{\mathbf{a}}_s)\}$ . Now let us expand  $\sigma_4(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_j \boxtimes \tilde{\mathbf{a}}_r \boxtimes \tilde{\mathbf{a}}_s)$ . Since the order of  $(i, j, r, s)$  does not matter, we will consider just the three cases

$$(i, j, r, s) \in \{(1, 1, 2, 2), (1, 1, 2, 3), (1, 2, 3, 4)\}.$$

Direct calculation shows that

$$\begin{aligned} \mathbf{Z}^{(1122)} &= \frac{1}{2} \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2)(\tilde{\mathbf{a}}_2)^{\top} + \frac{1}{2} \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2)(\tilde{\mathbf{a}}_1)^{\top} \\ &= \frac{1}{2} \begin{bmatrix} \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2) & \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2) \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_2 \end{bmatrix}^{\top}, \end{aligned}$$

hence  $\text{range}\{\mathbf{Z}^{(1122)}\} = \text{Span}\{\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2, \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_2\}$ . Similarly

$$\mathbf{Z}^{(1123)} = \frac{1}{4} \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2)(\tilde{\mathbf{a}}_3)^{\top} + \frac{1}{4} \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_3)(\tilde{\mathbf{a}}_2)^{\top} + \frac{1}{2} \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_3)(\tilde{\mathbf{a}}_1)^{\top},$$

hence  $\text{range}\{\mathbf{Z}^{(1123)}\} = \text{Span}\{\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2, \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_3, \tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_3\}$ . Finally,

$$\mathbf{Z}^{(1234)} = \frac{1}{4} \begin{bmatrix} \sigma_3(\tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_3 \boxtimes \tilde{\mathbf{a}}_4) & \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_3 \boxtimes \tilde{\mathbf{a}}_4) & \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_4) & \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_3) \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{a}}_1 & \tilde{\mathbf{a}}_2 & \tilde{\mathbf{a}}_3 & \tilde{\mathbf{a}}_4 \end{bmatrix}^{\top},$$

thus  $\text{range}\{\mathbf{Z}^{(1234)}\} = \text{Span}\{\sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_3), \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_4), \sigma_3(\tilde{\mathbf{a}}_1 \boxtimes \tilde{\mathbf{a}}_3 \boxtimes \tilde{\mathbf{a}}_4), \sigma_3(\tilde{\mathbf{a}}_2 \boxtimes \tilde{\mathbf{a}}_3 \boxtimes \tilde{\mathbf{a}}_4)\}$ .

Thus, we have that

$$(10.7) \quad \text{range}\{\mathbf{P}_{\text{sym}}\} \subseteq \text{Span}\{\{\sigma_3(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_j \boxtimes \tilde{\mathbf{a}}_r) : 1 \leq i < r \leq R, 1 \leq j \leq R\}\},$$

Assume that  $\text{krank}\{\mathbf{F}\} = \text{krank}\{\mathbf{F}\}^\top \geq 2$ . Fix  $1 \leq i < r \leq R$  and  $1 \leq j \leq R$ , the submatrix  $\mathbf{F}_{(i,r),:}$  has rank 2. Note that the  $j$ -th column  $\mathbf{F}_{(i,r),j}$  of that matrix is nonzero (by [Assumption 2](#)), and therefore there must exist at least one other column  $\mathbf{F}_{(i,r),s}$  that is not a multiple of  $\mathbf{F}_{(i,r),j}$ , which implies  $\det \mathbf{F}_{(i,r),(j,s)} \neq 0$ . Therefore, in view of (10.6), we have  $\sigma_3(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_r \boxtimes \tilde{\mathbf{a}}_j)$ , and hence the two spaces in (10.7) are equal due to the dimension count (which also implies  $\dim \text{range}\{\mathbf{P}_{\text{sym}}\} = \dim(S^3(\mathbb{R}^R)) - R$ ).

Now assume that  $\text{krank}\{\mathbf{F}\} = 1$ . This means that there exists  $1 \leq k < \ell \leq R$  such that the submatrix  $\mathbf{F}_{(k,\ell),:}$  is rank 1. In particular,  $\det \mathbf{F}_{(k,\ell),(k,m)} = 0$  for any  $m$ . This implies that

$$\text{range}\{\mathbf{P}_{\text{sym}}\} \subseteq \text{Span}\{\{\sigma_3(\tilde{\mathbf{a}}_i \boxtimes \tilde{\mathbf{a}}_j \boxtimes \tilde{\mathbf{a}}_r) : 1 \leq i < r \leq R, 1 \leq j \leq R, (i, r, j) \neq (k, \ell, k)\}\},$$

and hence the two linear spaces are not equal in (10.7), and thus  $\dim \text{range}\{\mathbf{P}_{\text{sym}}\} < \dim(S^3(\mathbb{R}^R)) - R$ .

Finally, we note that in case when equality holds in (10.7),

$$(\mathbf{a}_k \boxtimes \mathbf{a}_k \boxtimes \mathbf{a}_k)^\top \mathbf{P}_{\text{sym}} = 0, \quad k = 1, \dots, R,$$

which completes the proof.  $\square$

**11. Numerical experiments.** A preliminary implementation is available in MATLAB at <https://github.com/kdu/paratuck-2>, together with the reproducible experiments. We use the following metrics to evaluate the results of the algorithms:

- squared Frobenius error:

$$MSE(\mathcal{T}, \hat{\mathcal{T}}) = \|\mathcal{T} - \hat{\mathcal{T}}\|_F^2$$

(to measure the reconstruction error);

- sum of squared sines of angles between columns (to measure the correct recovery of the factors)

$$SSS(\mathbf{A}, \hat{\mathbf{A}}) = \sum_{k=1}^R 1 - \left( \frac{\langle \mathbf{a}_k, \hat{\mathbf{a}}_k \rangle}{\|\mathbf{a}_k\| \|\hat{\mathbf{a}}_k\|} \right)^2$$

where the columns of  $\hat{\mathbf{A}}$  are permuted to match the columns of  $\mathbf{A}$  beforehand. The permutation is computed based on Hungarian algorithm. Note that  $SSS(\mathbf{A}, \hat{\mathbf{A}}) = \frac{\|\mathbf{A} - \hat{\mathbf{A}}\|_F^2}{2}$  if the columns of  $\mathbf{A}$  and  $\hat{\mathbf{A}}$  are normalized.

We note that in the implementation we use a basic joint eigenvalue computation algorithm, and thus the performance can be improved by using generalized Schur decompositions or approximate joint diagonalization techniques.

**11.1. Nonsymmetric case: simplest example.** We consider the following simple test example:

$$\mathbf{F} = \begin{bmatrix} 1 & 1 \\ 2 & -1 \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} 1 & 1 \\ 1 & -3 \end{bmatrix},$$

$$\mathbf{G} = \begin{bmatrix} -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ 1 & 0 & 2 & 1 & -3 & 2 & -2 & -1 & 0 & 1 \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

The squared Frobenius norm of  $\mathcal{T} = \text{PT2D}(\mathbf{A}, \mathbf{B}, \mathbf{F}, \mathbf{G}, \mathbf{H})$  is 13538.

Algorithm 5.1 returns a Paratuck-2 approximation with approximation error as  $MSE(\mathcal{T}, \hat{\mathcal{T}}) = 1.7 \cdot 10^{-26}$  and factor distances  $SSS(\mathbf{A}, \hat{\mathbf{A}}) = 6.6 \cdot 10^{-26}$  and  $SSS(\mathbf{B}, \hat{\mathbf{B}}) = 2.2 \cdot 10^{-26}$ .

We also ran ALS(alternating least squares) [2, pp. 68–71] with random initialization (factors drawn from the uniform distribution). With 5000 iterations, ALS gave an approximation error  $MSE(\mathcal{T}, \hat{\mathcal{T}})$  below  $10^{-10}$  only in 3% of the cases (number of random initializations was 100).

**11.2. Random example.** We next consider random example with  $I = J = 10$ ,  $R = S = 4$  and  $K = 100$ . We generate  $\mathbf{A}$ ,  $\mathbf{B}$ ,  $\mathbf{H}$ ,  $\mathbf{G}$  having i.i.d. entries drawn randomly from  $N(0, 1)$ , and  $\mathbf{F}$  from uniform distribution in  $[0.5, 1.5]$  (to guarantee [Assumptions 2](#) and [7](#).) without any perturbation explicitly added to the tensor. For all 1000 realizations of the factors, [Algorithm 5.1](#) was able to accurately compute the factors of the PT2D, as shown in Fig. 11.2.

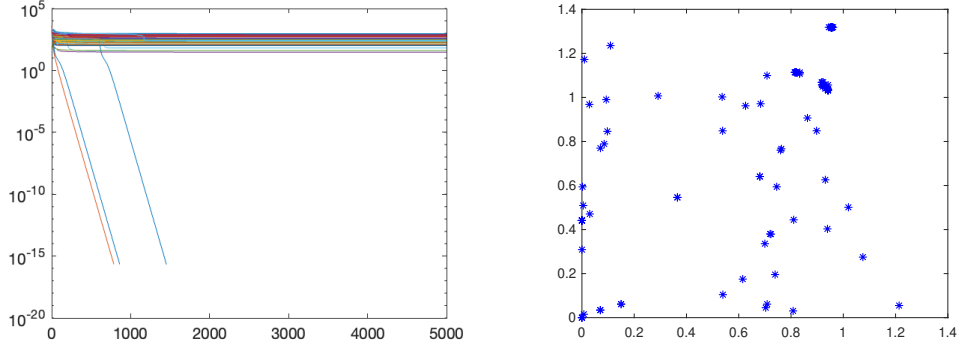


Fig. 11.1: Performance of the plain ALS for the  $2 \times 2$  example. Left: convergence plots for 100 initializations. Right:  $SSS(\mathbf{A}, \hat{\mathbf{A}})$  versus  $SSS(\mathbf{B}, \hat{\mathbf{B}})$  after 5000 ALS iterations.

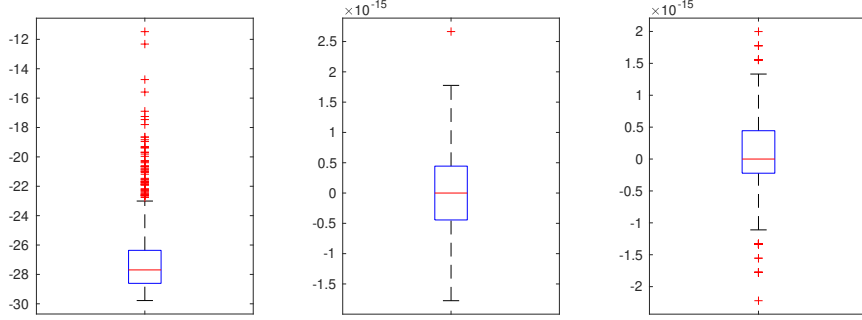


Fig. 11.2: Performance of Algorithm 5.1 for random examples with  $(R, S) = (4, 4)$ , boxplots for 1000 realizations. Left: relative error in logarithmic scale  $\log_{10}(MSE(\mathcal{T}, \hat{\mathcal{T}})/\|\mathcal{T}\|^2)$ , middle:  $SSS(\mathbf{A}, \hat{\mathbf{A}})$ . Right:  $SSS(\mathbf{B}, \hat{\mathbf{B}})$ .

*Noisy scenario.* We also pick a single realization of factors from the previous experiments, and test the quality of the reconstruction of the factors for the case of additive noise  $\mathcal{T} = \mathcal{T}_0 + \mathcal{E}$ . The level of noise is measured in dB, where  $X$  dB means that  $10 \log_{10}(\|\mathcal{T}_0\|_F^2 / \|\mathcal{E}\|_F^2) = X$ . For example, 40dB corresponds to the case  $\|\mathcal{E}\|_F = 10^{-2} \|\mathcal{T}_0\|_F$ . In Fig. 11.3, we show the simulation results for random noise from  $N(0, 1)$ , normalized to have achieve a prescribed noise level (essentially uniform distribution on a high-dimensional hypersphere  $\|\mathcal{E}\|_F = C$ ).

The results in Fig. 11.3 show that Algorithm 5.1 still works for small perturbations, but the performance degrades for higher levels of noise. We have tried an improvement based on taking into account the structure of the subspace spanned by matrices  $\mathbf{P}_k = \pi(\mathbf{p}_k)$  from Theorem 5.4. Note that from Proposition 9.1 it follows that the (noiseless)  $R^2 \times S^2 \times M$  tensor  $\mathcal{P}$  build as  $\mathcal{P}_{::,k} = \mathbf{P}_k$ , must have the multilinear ranks equal to  $((\binom{R}{2}, \binom{S}{2}), \cdot)$ . Therefore, before performing steps 5 of Algorithm 5.1, we can compute the Tucker-2 approximation of  $\mathcal{P}$  by the higher-order orthogonal iteration (HOOI). In our experiments (Fig. 11.3) HOOI gives a visible improvement, and we only need a few iterations of HOOI.

*Singular values.* However, there is a fundamental limitation of Algorithm 5.1, which is illustrated by the plots of singular values in Fig. 11.4. We see that  $\Phi(\mathcal{T})$  has  $\binom{RS+1}{2}$  singular values in total, out of which we are looking for  $\binom{R}{2} \binom{S}{2}$  smallest singular values. When  $\mathcal{T}$  is perturbed by the additive noise, this also



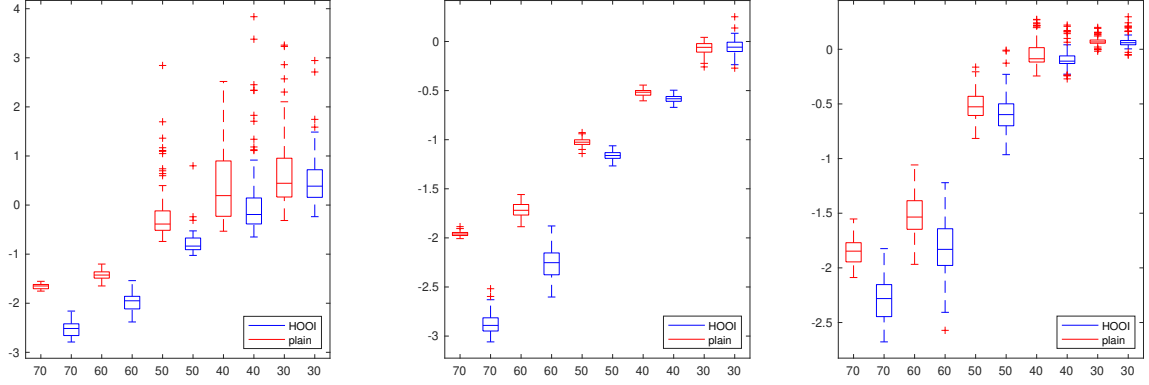


Fig. 11.3: Performance of Algorithm 5.1 with respect to additive noise, for 100 realizations (plots in logarithmic scale for easier visualization). Left: relative error  $\log_{10}(MSE(\mathcal{T}, \hat{\mathcal{T}})/\|\mathcal{T}\|^2)$ . Middle:  $\log_{10}(SSS(\mathbf{A}, \hat{\mathbf{A}}))$ . Right:  $\log_{10}(SSS(\mathbf{B}, \hat{\mathbf{B}}))$ . For each of the dB levels, the pair of columns is given: without and with HOOI improvement (10 iterations).

results in a nonlinear perturbation of  $\Phi(\mathcal{T})$ .

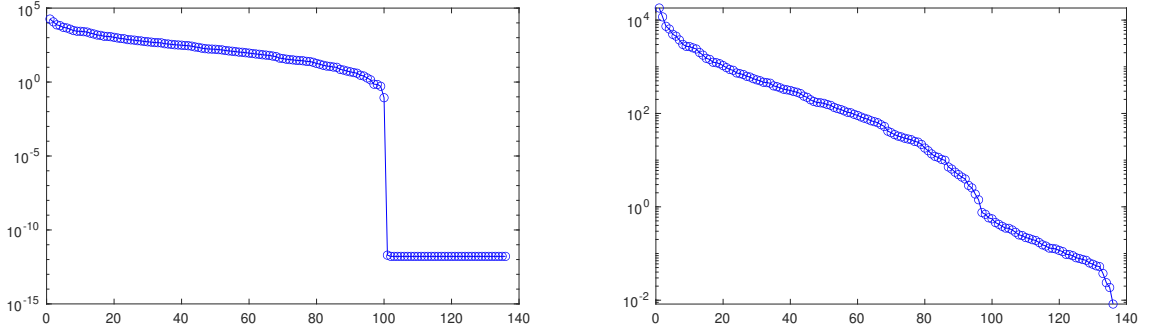


Fig. 11.4: Singular values of  $\Phi(\mathcal{T})$ . Left: noiseless scenario. Right: noise level 60dB.

**11.3. Symmetric case.** Finally, to illustrate the performance of the algorithm for the DEDICOM (symmetric PT2D), we consider the example of PARAFAC-2 decomposition as in Example 3.5. We take  $I = 30$ ,  $J_k = 40$ ,  $K = 20$  and  $R = 3$ , where the factors  $\mathbf{A}$  is generated randomly from the uniform distribution in  $[1, 2]$  and  $\mathbf{B}_k$  (of the same size),  $\mathbf{D}_k$  are with elements drawn from the uniform distribution in  $[0, 1]$ . This setup was taken from the manual of the TensorLy package [20], and we generate a single example, stacking all the Note that in this case the columns of the matrix  $\mathbf{A} \in \mathbb{R}^{30 \times 3}$  are highly correlated, with the correlation matrix of the columns given by

$$\approx \begin{bmatrix} 1 & 0.97 & 0.95 \\ 0.97 & 1 & 0.96 \\ 0.95 & 0.96 & 1 \end{bmatrix}.$$

We proceed by adding additive noise  $\mathcal{T} = \mathcal{T}_0 + \mathcal{E}$  for different noise level and use the method in Example 3.5 to reduce to the DEDICOM decomposition problem, with dimension reduction using the SVD of the first unfolding of  $\mathcal{T}$ :

- the SVD of the first unfolding  $\mathbf{T}^{(1)} = \tilde{\mathbf{U}}\tilde{\Sigma}\tilde{\mathbf{V}}^T$  is computed;
- the matrices  $\mathbf{X}_k$  are projected on the leading  $R$  left singular vectors:  $\tilde{\mathbf{X}}_k = (\Sigma_{1:R, 1:R})^{-1}(\mathbf{U}_{1:R, :}^T)\mathbf{X}_k$ .

The algorithm [Algorithm 6.1](#) is then used to compute the DEDICOM approximation of a tensor  $\widetilde{\mathcal{M}} \in \mathbb{R}^{3 \times 3 \times 20}$  with slices  $\widetilde{\mathcal{M}}_{:,:,k} = \widetilde{\mathbf{X}}_k(\widetilde{\mathbf{X}}_k)^\top$ . For reference, we plot in [Figure 11.5](#) the singular values of the matrix  $\Phi(\widetilde{\mathcal{M}})$ .

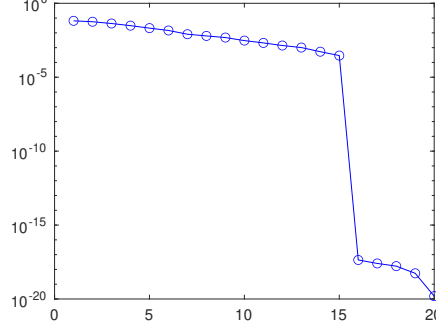


Fig. 11.5: Singular values of  $\Phi(\widetilde{\mathcal{M}})$ .

In [Figure 11.6](#), we report the results of approximation computed by DEDICOM. We see that the algorithm gives low error on the difference between the true and estimated factor  $\mathbf{A}$ . We also plot the relative errors for estimation of the true (uncompressed)  $\mathcal{M} \in \mathbb{R}^{I \times I \times K}$ , and see that the algorithm works for higher levels of noise than in the nonsymmetric case.

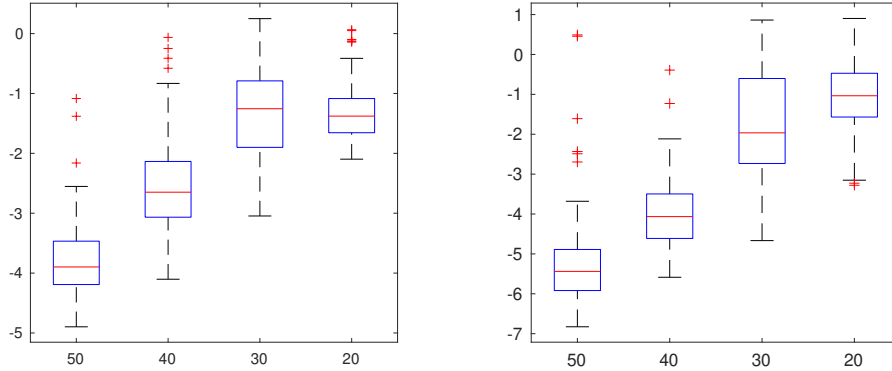


Fig. 11.6: Reconstructions of the factor and of the tensor, for different noise levels (in dB), for 100 noise realizations. Left: box plots of  $\log_{10}(SSS(\mathbf{A}, \widehat{\mathbf{A}}))$ . Right: relative errors  $\log_{10}(MSE(\mathcal{M}, \widehat{\mathcal{M}})/\|\mathcal{M}\|^2)$ .

In terms of computational speed, the whole computation of a compression of matrices  $\mathbf{X}_k$  together with DEDICOM took just a few milliseconds in MATLAB R2024a on a MacBook laptop with M3 Pro chipset. As a comparison, running the PARAFAC-2 decomposition for the same example with an iterative algorithm from [\[20\]](#) took  $\approx 5 - 6$  seconds for 500 iterations.

**Conclusion.** In this paper we proposed a lifting approach to ParaTuck-2 decomposition and DEDICOM. This approach clarifies and relaxes the existing uniqueness conditions and leads to simple algebraic algorithms. The algebraic algorithms, that use just nullspace and eigenvalue computations, are the first (except the case of  $R = 2$  in DEDICOM), up to the author's knowledge, that are guaranteed to compute the

exact decomposition and are also applicable in the approximation scenario, as shown by the numerical experiments.

The main drawback of the approach is its computational complexity, as it requires finding an SVD of a matrix with  $\binom{R+1}{2}\binom{S+1}{2}$  matrix, thus the complexity grows very quickly as  $R, S$  increase. Another issue is that we need  $\binom{R}{2}\binom{S}{2}$  last singular vectors, a number that also grows with  $R, S$  and can affect the robustness to noise and numerical stability of the algorithm for high  $R, S$ . However, we believe, that there are plenty of possible improvements of the basic algebraic algorithms presented in this paper, and that lifting approach can serve as the base for reliable algorithms to compute ParaTuck-2 decompositions.

**Acknowledgements.** I would like to thank many colleagues for stimulating discussions.

## REFERENCES

- [1] B. W. BADER, R. A. HARSHMAN, AND T. G. KOLDA, *Temporal analysis of semantic graphs using ASALSAN*, in Seventh IEEE international conference on data mining (ICDM 2007), IEEE, 2007, pp. 33–42.
- [2] R. BRO, *Multi-way Analysis in the Food Industry*, PhD thesis, Vrije Universiteit Brussel (VUB), 1998.
- [3] A. L. F. DE ALMEIDA, G. FAVIER, AND L. R. XIMENES, *Space-time-frequency (STF) MIMO communication systems with blind receiver based on a generalized PARATUCK2 model*, IEEE Transactions on Signal Processing, 61 (2013), pp. 1895–1909.
- [4] J. DE JONGHE, K. USEVICH, P. DREESEN, AND M. ISHTEVA, *Compressing neural networks with two-layer decoupling*, in 2023 IEEE 9th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), 2023, pp. 226–230, <https://doi.org/10.1109/CAMSAP58249.2023.10403509>.
- [5] P. M. R. DE OLIVEIRA, C. A. R. FERNANDES, G. FAVIER, AND R. BOYER, *PARATUCK Semi-Blind Receivers for Relaying Multi-Hop MIMO Systems*, Digital Signal Processing, 92 (2019), pp. 127–138, <https://doi.org/10.1016/j.dsp.2019.05.011>.
- [6] P. DIACONIS AND B. STURMFELS, *Algebraic algorithms for sampling from conditional distributions*, The Annals of Statistics, 26 (1998), pp. 363 – 397, <https://doi.org/10.1214/aos/1030563990>.
- [7] I. DOMANOV AND L. D. LATHAUWER, *Canonical polyadic decomposition of third-order tensors: Reduction to generalized eigenvalue decomposition*, SIAM Journal on Matrix Analysis and Applications, 35 (2014), pp. 636–660, <https://doi.org/10.1137/130916084>.
- [8] P. DREESEN, M. ISHTEVA, AND J. SCHOUKENS, *Decoupling multivariate polynomials using first-order information and tensor decompositions*, SIAM Journal on Matrix Analysis and Applications, 36 (2015), pp. 864–879, <https://doi.org/10.1137/140991546>.
- [9] G. FAVIER AND A. L. DE ALMEIDA, *Overview of constrained PARAFAC models*, EURASIP Journal on Advances in Signal Processing, 2014 (2014), p. 142.
- [10] L. D. GARCIA, M. STILLMAN, AND B. STURMFELS, *Algebraic geometry of Bayesian networks*, Journal of Symbolic Computation, 39 (2005), pp. 331–355, <https://doi.org/10.1016/j.jsc.2004.11.007>. Special issue on the occasion of MEGA 2003.
- [11] R. A. HARSHMAN, *Models for analysis of asymmetrical relationships among  $n$  objects or stimuli*, in First Joint Meeting of the Psychometric Society and the Society for Mathematical Psychology, McMaster University, Hamilton, Ontario, Aug. 1978.
- [12] R. A. HARSHMAN AND M. E. LUNDY, *Uniqueness proof for a family of models sharing features of Tucker’s three-mode factor analysis and parafac/candecomp*, Psychometrika, 61 (1996), pp. 133–154.
- [13] H. A. KIERS, *An alternating least squares algorithm for fitting the two-and three-way DEDICOM model and the idioscal model*, Psychometrika, 54 (1989), pp. 515–521.
- [14] H. A. KIERS, *An alternating least squares algorithm for PARAFAC2 and three-way DEDICOM*, Computational Statistics & Data Analysis, 16 (1993), pp. 103–118, [https://doi.org/10.1016/0167-9473\(93\)90247-Q](https://doi.org/10.1016/0167-9473(93)90247-Q).
- [15] H. A. KIERS, J. M. TEN BERGE, AND R. BRO, *PARAFAC2—Part I. a direct fitting algorithm for the PARAFAC2 model*, Journal of Chemometrics: A Journal of the Chemometrics Society, 13 (1999), pp. 275–294.
- [16] T. G. KOLDA AND B. W. BADER, *Tensor decompositions and applications*, SIAM review, 51 (2009), pp. 455–500.
- [17] K. NASKOVSKA, *Advanced tensor based signal processing techniques for wireless communication systems and biomedical signal processing*, PhD thesis, Ilmenau, Jan 2020. Dissertation, Technische Universität Ilmenau, 2019.
- [18] M. ROALD, C. SCHENKER, V. CALHOUN, T. ADALI, R. BRO, J. COHEN, AND E. ATAMAN, *An AO-ADMM approach to constraining PARAFAC2 on all modes*, SIAM Journal on Mathematics of Data Science, 4 (2022), pp. 1191–1222, <https://doi.org/10.1137/21M1450033>.
- [19] J. M. TEN BERGE AND H. A. KIERS, *Some uniqueness results for PARAFAC2*, Psychometrika, 61 (1996), pp. 123–132.
- [20] TENSORLY DEVELOPERS, *Demonstration of parafac2 (tensorly package manual)*, 2024. [tensorly.org/stable/auto\\_examples/decomposition/plot\\_parafac2.html](https://tensorly.org/stable/auto_examples/decomposition/plot_parafac2.html) [Accessed: 13 November 2024].
- [21] K. USEVICH, Y. ZNIYED, M. ISHTEVA, P. DREESEN, AND A. DE ALMEIDA, *Tensor-based two-layer decoupling of multivariate polynomial maps*, in 31st European Signal Processing Conference, EUSIPCO 2023, Helsinki, Finland, Sept. 2023, European Association for Signal Processing, <https://doi.org/10.23919/EUSIPCO58844.2023.10289900>.
- [22] Y. ZNIYED AND A. L. DE ALMEIDA, *A stochastic algorithm for the ParaTuck decomposition*, Digital Signal Processing, 156 (2025), p. 104767.