



HAL
open science

A Generalization to Schur's Lemma with an Application to Joint Independent Subspace Analysis

Dana Lahat, Christian Jutten

► **To cite this version:**

Dana Lahat, Christian Jutten. A Generalization to Schur's Lemma with an Application to Joint Independent Subspace Analysis. 2015. hal-01247899v1

HAL Id: hal-01247899

<https://hal.science/hal-01247899v1>

Preprint submitted on 23 Dec 2015 (v1), last revised 6 Jun 2016 (v2)

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.

A Generalization to Schur's Lemma with an Application to Joint Independent Subspace Analysis

Dana Lahat and Christian Jutten, *Fellow, IEEE*[†]

Abstract

This paper has a threefold contribution. First, it introduces a generalization to Schur's lemma from 1905 on irreducible representations. Second, it provides a comprehensive uniqueness analysis of a recently-introduced source separation model. Third, it reinforces the link between signal processing and representation theory, a field of algebra that is more often associated with quantum mechanics than with signal processing. The source separation model that this paper relies on performs joint independent subspace analysis (JISA) using second order statistics. In previous work, we derived the Fisher information matrix (FIM) that corresponds to this model. The uniqueness analysis in this paper is based on analysing the FIM, and the derivation is based on our proposed generalization to Schur's lemma. We provide proof both to the new lemma and to the uniqueness conditions. From a different perspective, the generalization to Schur's lemma is inspired by a coupled matrix block diagonalization problem that arises from the JISA model. The results in this paper generalize previous results about identifiability of independent vector analysis (IVA) using second order statistics. The results in this paper complement previously-known results on the uniqueness of joint block diagonalization (JBD) and block term decompositions (BTD), as well as of their coupled counterparts.

Keywords Uniqueness, blind source separation, independent subspace analysis, independent vector analysis, data fusion, irreducible representations, coupled decompositions, joint block diagonalization

1 Introduction

A well-known result in algebra, group theory and irreducible representations ([1, 2, 3, 4] and others) is often known as “Schur's lemma” [5].

In previous work [6], Schur's lemma arose naturally in the analysis of the uniqueness of blind source separation (BSS) of an instantaneous invertible mixture of piecewise stationary real multidimensional sources that using second-order statistics (SOS). This BSS model can be reformulated as a symmetric joint block diagonalization (JBD) of a set of covariance matrices. The analysis in [6] boiled down to showing that non-uniqueness existed for irreducible data if only if at least two multidimensional sources exist in equivalent subspaces. This result is complementary to the generic uniqueness analysis in [7]. In [7], it was shown that JBD was a special case of a more general tensor decomposition, block term decomposition (BTD). The generic uniqueness conditions for JBD/BTD were derived, and it was stated that “In the nongeneric case, lack of uniqueness can be due to the fact that tensors can be further block-diagonalized”, or “be subdivided in smaller blocks”. Our result in [6] characterises the cases where the model is non-identifiable in the nongeneric case and when the blocks are irreducible, that is, cannot be further divided into smaller blocks. The analysis in [6] is based on characterising the non-invertibility of the Fisher information matrix (FIM).

In this paper, we follow a similar analytical approach. Based on previous derivation of the FIM for a different BSS model, called joint independent subspace analysis (JISA) [8, 9], we characterise the points of non-invertibility of this FIM. Interestingly, this resulted in the need to use a different lemma on irreducible subspaces. This new lemma, which we present in this paper, can be regarded as a generalization of Schur's original lemma. In analogy to the results in [6], also in this case the non-identifiability conditions for the irreducible case (i.e. nongeneric uniqueness) can be stated as an equivalence of subspaces of a pair of sources. However, in this case, the equivalence is in a generalized sense, as we define in this paper.

The results in [6], as well as in this current paper, conform with those in [7] (and references therein) about invariant subspaces.

JISA is a generalization of independent vector analysis (IVA) to multidimensional components, or subspaces. Accordingly, the uniqueness results in this paper generalize those in [10], for IVA, to the multidimensional

^{*}D. Lahat and Ch. Jutten are with GIPSA-Lab, UMR CNRS 5216, Grenoble Campus, BP46, F-38402 Saint Martin d'Hères, France. email: {Dana.Lahat, Christian.Jutten}@gipsa-lab.grenoble-inp.fr.

[†]This work is supported by the project CHESS, 2012-ERC-AdG-320684. GIPSA-Lab is a partner of the LabEx PERSYVAL-Lab (ANR-11-LABX-0025).

case. This is in analogy to the results in [6], which generalize the well-known “spectral diversity” condition for one-dimensional BSS with temporal or spectral diversity to multidimensional subspaces.

The interpretation of the results in this paper, in terms of signal processing and data fusion, follows similar lines as its IVA counterpart in [10] and references therein.

We mention that in the IVA or non-stationary one-dimensional BSS case there is no need to resort to Schur’s lemma (or a variant thereof), because the blocks are scalar; hence, there is no irreducibility issue: the blocks in the algebraic formulation of the model, be it joint diagonalization (JD) or coupled JD [11, 12], are already of size one, and thus cannot be further reduced. For this reason, the derivation of uniqueness of IVA in [10] does not involve Schur’s lemma despite the fact that it follows the same methodology as in this paper.

In this paper, we chose, for simplicity and clarity of exposition, to deal with a simple instance of JISA. However, more elaborate formulations exist (e.g. [13]) that allow variability between data sets, for example in the size of the blocks, as well as mixing matrices of different sizes, possibly rectangular. We assume real data but our results can be readily generalized to the complex domain. In particular, one can join the non-stationarity diversity of [6] with the multiset diversity of JISA, as a natural generalization of the model in e.g. [11, 12]. The model in [11, 12] amounts to a coupled JD, whereas in our case, the model would amount to coupled JBD. One can combine both results of uniqueness, [6] and of this paper, and conclude that for coupled JBD, non-uniqueness exists if and only if one can find equivalence of subspaces both in the data set and in the temporal or frequency coordinates. In particular, JBD/BTD has strong uniqueness also in the underdetermined case [7]. As a result, these properties are inherited to the coupled formulation and further reinforce its uniqueness. Roughly speaking, the larger the number of types of diversity in the model, the smaller the risk to fall randomly into a situation of non-uniqueness.

Our interest in the case of SOS-based real-valued JISA with only data set diversity and no spatio/temporal diversity (as in e.g. [11, 12, 14]) stems from the fact that this is the only case in which there is absolutely no identifiability of each data set alone. Hence, it is the most challenging scenario, and illustrates the power of SOS-based JISA. The comprehension that coupled factorizations may be unique even in cases where individual factorizations are not is not new, see e.g. [15], [16] and references therein.

The link between irreducible representations and JBD in the context of signal processing was first introduced, to the best of our knowledge, by [17, 18, 19, 20]. The main difference of this work from these (and from [6]) is that in this work we provide a substantial generalization to well-known results in algebra, and this is achieved by analysing a new type of a BSS model as well as a new type of coupled factorization.

The main novelty and contribution of this paper is the lemmas in Section 3 and the uniqueness Theorem 4.1. Some of the results on partitioned matrices, in Appendix A, are also new.

Part of this work was presented in [21].

The notations in this paper follow those of [9], unless stated otherwise.

The rest of this paper is as follows. In Section 2 we introduce some operations on sets of matrices that will be used throughout this paper. Section 3 presents our analogue to Schur’s lemma. In Section 4, we apply the new lemma to the uniqueness analysis of a basic form of JISA. We briefly discuss our results in Section 5.

2 Basic Definitions

The following definitions that will be used throughout this paper.

2.1 Admissible Data

Data is non-admissible if it breaks the model assumptions on which our analysis is based, or it is physically implausible in various senses. Conversely, the whole identifiability analysis is valid only for data for which these assumptions hold. We make a distinction between two types of admissible data: “strict”, and “relaxed”. The “strict” form corresponds to commutation lemmas more similar to Schur’s original lemma, whereas “relaxed” admits a broader range of scenarios that are plausible in the BSS/JISA sense. The “strict” sense corresponds to \mathbf{S}_{ii} with no zeros, i.e. sources that are mutually dependent across all mixtures, whereas “relaxed”, as its name implies, relaxes this assumption.

Definition 2.1 (Admissible and non-admissible \mathbf{P} and \mathbf{R}). Consider a $KP \times KP$ matrix \mathbf{P} , $\mathbf{P}^{[k,l]}$ its (k, l) th $P \times P$ block, $k, l = 1, \dots, K$. In order to be admissible, \mathbf{P} must have the following properties; either in the strict or relaxed sense. Otherwise, it is non-admissible.

1. \mathbf{P} is a positive definite covariance matrix. Hence,
 - (a) \mathbf{P} invertible

(b) \mathbf{P} symmetric s.t. $\mathbf{P}^{[k,l]} = \mathbf{P}^{[l,k]\top} \forall k, l$

2. $\mathbf{P}^{[k,l]}$ must reflect covariance between true multidimensional components. Hence,

• Strict:

(a) $\mathbf{P}^{[k,l]} = \mathbf{P}^{[l,k]\top}$ invertible $\forall k, l$

(b) the set $\{\mathbf{P}^{[k,l]}\}_{k,l=1}^K$ irreducible (in the sense of Definition 2.2 on page 3)

• Relaxed:

(a) either $\mathbf{P}^{[k,l]} = \mathbf{0}$ or $\mathbf{P}^{[k,l]}$ invertible and has no zero entries, for each $k \neq l$

(b) let us collect all indices k for which $\mathbf{P}^{[k,l]} \neq \mathbf{0}$ in $\mathcal{D} = \{1, \dots, D\}$. Then, the set $\{\mathbf{P}^{[k,l]}\}_{\{k,l\} \in \mathcal{D}}$ is irreducible (in the sense of Definition 2.2 on page 3)

Note: due to symmetry, we need to consider only $k \leq l$.

2.2 Generalized Properties

In this section, we extend well-known operations from one to several sets of matrices, where these sets are coupled in a specific way that corresponds to the JISA model [8, 9].

Definition 2.2 (Irreducibility (in the generalized sense)). Consider a set of matrices $\mathbf{P}^{[k,l]}$, $k, l = 1, \dots, K$, all of dimension $P \times P$, where $\mathbf{P}^{[k,l]} = (\mathbf{P}^{[l,k]})^\top$ (sub-blocks of a symmetric $\mathbf{P} \in \mathbb{R}^{K \times K \times P}$).

The set is reducible in the generalized sense if there exist invertible matrices $\mathbf{T}^{[k]}$, $k = 1, \dots, K$, such that all

$$\mathbf{T}^{[k]} \mathbf{P}^{[k,l]} (\mathbf{T}^{[l]})^\top$$

are block diagonal with the same form. Otherwise, the set is irreducible (in the generalized sense).

• For $K \leq 2$, this set can always be exactly diagonalized (unless the matrices are defective), using generalized eigenvalue decomposition (GEVD) [22, Chapter 12.2, Equation (53)]; hence, this condition applies only to $K \geq 3$

• Due to symmetry, we may consider only $k \leq l$.

Definition 2.3 (Generalized similarity transformation (or equivalence)). Two sets of matrices, $\{\mathbf{R}^{[k,l]}\}_{k,l=1}^K \in \mathbb{R}^{R \times R}$ and $\{\mathbf{P}^{[k,l]}\}_{k,l=1}^K \in \mathbb{R}^{P \times P}$ are said to be related by a generalized similarity transformation if

$$\mathbf{P}^{[k,l]} = \mathbf{L}^{[k]} \mathbf{R}^{[k,l]} \mathbf{L}^{-[l]}, \quad \forall k, l$$

for some K invertible matrices $\{\mathbf{L}^{[k]}\}_{k=1}^K$.

3 A Multiset Analogue to Schur's Lemma

In the first part of this paper, we present a multiset extension to Schur's Lemma [5].

Lemma 1 (Multiset analogue to Schur's first lemma). *Let \mathcal{P} be an irreducible (in the generalized sense) set of invertible $P \times P$ matrices $\mathbf{R}^{[k,l]}$, $k, l = 1, \dots, K$, where $\mathbf{P}^{[k,l]} = (\mathbf{P}^{[l,k]})^\top \forall k, l$. Let*

$$\mathbf{L}^{[k]} \mathbf{R}^{[k,l]} = \mathbf{R}^{[k,l]} \mathbf{L}^{[l]} \quad \forall k, l$$

Then, $\mathbf{L}^{[k]} = \mu \mathbf{I}_P \forall k$, $\mu \geq 0$.

Note: due to symmetry, we can work only on $k \leq l$.

Proof of Lemma 1. The proof is very similar to that of Schur's original lemma, see e.g. [4, Chapter 4], where the symmetry of the matrices replaces the original requirement that they form a representation. \square

Lemma 2 (Multiset analogue to Schur's second lemma, strict). *Let \mathcal{R}, \mathcal{P} be two irreducible (in the generalized sense) sets of invertible matrices $\mathbf{R}^{[k,l]}, \mathbf{P}^{[k,l]}$, $k, l = 1, \dots, K$, admissible in the strict sense of Definition 2.1, of dimensions $R \times R$ and $P \times P$, respectively, $\mathbf{R}^{[k,l]} = (\mathbf{R}^{[l,k]})^\top$, $\mathbf{P}^{[k,l]} = (\mathbf{P}^{[l,k]})^\top$. Let*

$$\mathbf{L}^{[k]} \mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]} \mathbf{L}^{[l]} \quad \forall k, l$$

Then, either $\mathbf{L}^{[k]} = \mathbf{0}_{P \times P} \forall k$ or $\mathbf{L}^{[k]} = \lambda \mathbf{O}^{[k]}$ ($\mathbf{O}^{[k]}$ orthogonal, $\lambda > 0$), i.e. $P = R$, and \mathcal{R}, \mathcal{P} are related by a generalized similarity (or equivalence) transformation $\mathbf{P}^{[k,l]} = \mathbf{O}^{[k]} \mathbf{R}^{[k,l]} \mathbf{O}^{-[l]} \forall k, l$.

Lemma 3 (Multiset analogue to Schur's second lemma, relaxed). *Let \mathcal{R}, \mathcal{P} be two irreducible (in the generalized sense) sets of invertible matrices $\mathbf{R}^{[k,l]}, \mathbf{P}^{[k,l]}$, $k, l = 1, \dots, K$, admissible in the relaxed sense of Definition 2.1, of dimensions $R \times R$ and $P \times P$, respectively, $\mathbf{R}^{[k,l]} = (\mathbf{R}^{[l,k]})^\top$, $\mathbf{P}^{[k,l]} = (\mathbf{P}^{[l,k]})^\top$. Let*

$$\mathbf{L}^{[k]} \mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]} \mathbf{L}^{[l]} \quad \forall k, l$$

Then, either

1. $\mathbf{L}^{[k]} = \mathbf{0}_{P \times P} \forall k$ (implicitly, $P = R$), or
2. there exist, without loss of generality, $k \in \mathcal{D}$, $\mathcal{D} = \{1, \dots, D\}$, $D \geq 2$, such that $\mathbf{L}^{[k]} = \lambda \mathbf{O}^{[k]}$ ($\mathbf{O}^{[k]}$ orthogonal, implicitly $\lambda > 0$ and $P = R$) and $\mathbf{R}^{[k,l]} = \mathbf{0} = \mathbf{P}^{[k,l]}$ for $k \in \mathcal{D}$ and $l \notin \mathcal{D}$ (due to symmetry, holds also for $k \notin \mathcal{D}$ and $l \in \mathcal{D}$), such that $\mathbf{R}^{[k \in \mathcal{D}, l \in \mathcal{D}]}$ and $\mathbf{P}^{[k \in \mathcal{D}, l \in \mathcal{D}]}$ are related by a generalized similarity (or equivalence) transformation $\mathbf{P}^{[k,l]} = \mathbf{O}^{[k]} \mathbf{R}^{[k,l]} \mathbf{O}^{-[l]}$ for $k, l \in \mathcal{D}$, or
3. there exists, without loss of generality, $k \in \mathcal{D}$, $\mathcal{D} = \{1, \dots, D\}$, $D \geq 1$, such that

$$\mathbf{P}^{[k,l]} = \begin{cases} \delta_{kl} \mathbf{I} & k, l \in \mathcal{D} \\ \mathbf{0} & k \in \mathcal{D}, l \notin \mathcal{D} \end{cases}$$

and the same for \mathbf{R} .

The proof of Lemma 2-3 is in Appendix B.

4 Uniqueness and Identifiability of JISA

We now turn to the second part of this paper, which is an application of Lemma 2-3 to the identifiability of JISA [8, 9]. JISA is a source separation model that at its simplest form can be reformulated as a coupled matrix block diagonalization. For the definition of the JISA model, see [9].

As in JBD [23], and in some similarity to the identifiability analysis of IVA in [24, 25], it is possible to study the identifiability of the model through the properties of the FIM. For each pair (i, j) , it has been shown in [9] that the FIM corresponds to the symmetric positive semi-definite $2Km_i m_j \times 2Km_i m_j$ matrix

$$\mathcal{H} = \begin{bmatrix} \mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1} & \mathbf{I}_K \otimes \mathcal{T}_{m_j, m_i} \\ \mathbf{I}_K \otimes \mathcal{T}_{m_i, m_j} & \mathbf{S}_{ii} \boxplus \mathbf{S}_{jj}^{-1} \end{bmatrix} \quad (1)$$

where

$$\mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1} = \begin{bmatrix} \mathbf{S}_{jj}^{[1,1]} \otimes [\mathbf{S}_{ii}^{-1}]_{11} & \cdots & \mathbf{S}_{jj}^{[1,K]} \otimes [\mathbf{S}_{ii}^{-1}]_{11} \\ \vdots & & \vdots \\ \mathbf{S}_{jj}^{[K,1]} \otimes [\mathbf{S}_{ii}^{-1}]_{K1} & \cdots & \mathbf{S}_{jj}^{[K,K]} \otimes [\mathbf{S}_{ii}^{-1}]_{KK} \end{bmatrix}$$

is a $Km_i m_j \times Km_i m_j$ matrix whose (k, l) th block is $\mathbf{S}_{jj}^{[k,l]} \otimes [\mathbf{S}_{ii}^{-1}]_{kl}$ and has dimensions $m_i m_j \times m_i m_j$. Hence, $\mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1}$ is a matrix partitioned into blocks according to $m_i m_j \mathbf{1}_K = \underbrace{[m_i m_j, \dots, m_i m_j]^\top}_{K \text{ times}}$, both in rows and columns. $\mathbf{S}_{ii}^{[k,l]}$ and $[\mathbf{S}_{ii}^{-1}]_{kl}$ are the (k, l) th blocks of \mathbf{S}_{ii} and \mathbf{S}_{ii}^{-1} , respectively, and have dimensions $m_i \times m_i$. The superscript notation $[k, l]$ is to remind the fact that in JISA context, $\mathbf{S}_{ii}^{[k,l]}$ is the covariance between sources

$\mathbf{s}_i^{[k]}$ and $\mathbf{s}_i^{[l]}$ in data sets k and l , respectively. Since inverting \mathbf{S}_{ii} mixes all data sets, we do not use this type of notation for \mathbf{S}_{ii}^{-1} . Therefore, $\mathbf{S}_{ii}^{[k,l]} \triangleq [\mathbf{S}_{ii}]_{kl}$. In (1) we introduce the commutation matrix $\mathcal{T}_{P,Q} \in \mathbb{R}^{PQ \times PQ}$, where $\text{vec}\{\mathbf{M}^\dagger\} = \mathcal{T}_{P,Q} \text{vec}\{\mathbf{M}\}$ for any $\mathbf{M} \in \mathbb{R}^{P \times Q}$ [26]. More properties of the commutation matrix can be found in Appendix A.

Matrix \mathcal{H} in (1) is always well-defined, since it is derived based on the assumption that \mathbf{S}_{ii} and \mathbf{S}_{jj} are invertible covariance matrices, hence positive definite symmetric. Conversely, the identifiability results that are based on the analysis of \mathcal{H} are valid only for the case that \mathbf{S}_{ii} and \mathbf{S}_{jj} are invertible. For further discussion about admissible data, see Section 2.1 and Definition 2.1.

For the purpose of our analysis, we introduce a simplified notation, in which $\Theta \triangleq \mathbf{S}_{jj}$ and $\Xi \triangleq \mathbf{S}_{ii}$. Then, Θ_{kl} and $[\Xi^{-1}]_{kl}$ are $m_j \times m_j$ and $m_i \times m_i$ matrices, representing the (k, l) th blocks of Θ and Ξ^{-1} , according to the partitions $m_j \mathbf{1}_K = \underbrace{[m_j, \dots, m_j]^\top}_{K \text{ times}}$ and $m_i \mathbf{1}_K = \underbrace{[m_i, \dots, m_i]^\top}_{K \text{ times}}$, respectively. Using this simplified notation, the matrix \mathcal{H} in (1) can now be rewritten as

$$\mathcal{H} = \begin{bmatrix} \Theta \boxplus \Xi^{-1} & \mathbf{I}_K \otimes \mathcal{T}_{m_j, m_i} \\ \mathbf{I}_K \otimes \mathcal{T}_{m_i, m_j} & \Xi \boxplus \Theta^{-1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \otimes \mathcal{T}_{m_i, m_j} \end{bmatrix} \underbrace{\begin{bmatrix} \Theta \boxplus \Xi^{-1} & \mathbf{I} \\ \mathbf{I} & \Theta^{-1} \boxplus \Xi \end{bmatrix}}_{\mathcal{H}'} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_K \otimes \mathcal{T}_{m_i, m_j}^\top \end{bmatrix} \quad (2)$$

where the factorization in the second step is due to Identity 1 on page 9. Therefore, identifiability consists in defining the sufficient and necessary conditions for the invertibility and thus positive-definiteness of

$$\mathcal{H}' \triangleq \begin{bmatrix} \Theta \boxplus \Xi^{-1} & \mathbf{I} \\ \mathbf{I} & \Theta^{-1} \boxplus \Xi \end{bmatrix} = \begin{bmatrix} \mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1} & \mathbf{I} \\ \mathbf{I} & \mathbf{S}_{jj}^{-1} \boxplus \mathbf{S}_{ii} \end{bmatrix}. \quad (3)$$

4.1 Analyzing \mathcal{H}

For \mathcal{H}' to be positive-definite, we require that for any vector $\mathbf{x} \in \mathbb{R}^{2K m_i m_j \times 1}$,

$$0 < \mathbf{x}^\top \mathcal{H}' \mathbf{x} = \mathbf{x}^\top \mathbf{V}^\top \underbrace{\mathbf{V} \mathbf{x}}_{\mathbf{v}} = \mathbf{v}^\top \mathbf{v}. \quad (4)$$

Conversely, for \mathcal{H}' to be non-positive-definite, there must exist some non-zero $\mathbf{x} \in \mathbb{R}^{2K m_i m_j \times 1}$ such that

$$0 = \mathbf{x}^\top \mathcal{H}' \mathbf{x} = \mathbf{x}^\top \mathbf{V}^\top \underbrace{\mathbf{V} \mathbf{x}}_{\mathbf{v}} = \mathbf{v}^\top \mathbf{v} = \sum_{\alpha=1}^{2K m_i m_j} |v_\alpha|^2 \Leftrightarrow v_\alpha = 0 \forall \alpha \Leftrightarrow \mathbf{V} \mathbf{x} = \mathbf{0}. \quad (5)$$

4.1.1 Factorizing \mathcal{H}

First, based on (4) and (5), we look for a meaningful factorization $\mathcal{H}' = \mathbf{V}^\top \mathbf{V}$. We propose the following.

$$\mathcal{H}' \stackrel{(3)}{=} \begin{bmatrix} \Theta \boxplus \Xi^{-1} & \mathbf{I} \\ \mathbf{I} & \Theta^{-1} \boxplus \Xi \end{bmatrix} \quad (6a)$$

$$\stackrel{(7)}{=} \begin{bmatrix} \Theta^{\frac{1}{2}} \Theta^{\frac{1}{2} \top} \boxplus \Xi^{-\frac{1}{2} \top} \Xi^{-\frac{1}{2}} & \Theta^{-\frac{1}{2} \top} \Theta^{\frac{1}{2} \top} \boxplus \Xi^{\frac{1}{2}} \Xi^{-\frac{1}{2}} \\ \Theta^{\frac{1}{2}} \Theta^{-\frac{1}{2}} \boxplus \Xi^{-\frac{1}{2} \top} \Xi^{\frac{1}{2} \top} & \Theta^{-\frac{1}{2} \top} \Theta^{-\frac{1}{2}} \boxplus \Xi^{\frac{1}{2}} \Xi^{\frac{1}{2} \top} \end{bmatrix} \quad (6b)$$

$$\stackrel{\text{Identity 5}}{=} \begin{bmatrix} (\Theta^{\frac{1}{2} \top} \boxplus \Xi^{-\frac{1}{2}})^\top (\Theta^{\frac{1}{2}} \boxplus \Xi^{-\frac{1}{2}}) & (\Theta^{\frac{1}{2} \top} \boxplus \Xi^{-\frac{1}{2}})^\top (\Theta^{-\frac{1}{2}} \boxplus \Xi^{\frac{1}{2}})^\top \\ (\Theta^{-\frac{1}{2}} \boxplus \Xi^{\frac{1}{2}})^\top (\Theta^{\frac{1}{2}} \boxplus \Xi^{-\frac{1}{2}}) & (\Theta^{-\frac{1}{2}} \boxplus \Xi^{\frac{1}{2}})^\top (\Theta^{-\frac{1}{2}} \boxplus \Xi^{\frac{1}{2}})^\top \end{bmatrix} \quad (6c)$$

$$= \underbrace{\begin{bmatrix} (\Theta^{\frac{1}{2} \top} \boxplus \Xi^{-\frac{1}{2}})^\top \\ (\Theta^{-\frac{1}{2}} \boxplus \Xi^{\frac{1}{2}})^\top \end{bmatrix}}_{\mathbf{V}^\top} \underbrace{\begin{bmatrix} \Theta^{\frac{1}{2}} \boxplus \Xi^{-\frac{1}{2}} & \Theta^{-\frac{1}{2}} \boxplus \Xi^{\frac{1}{2}} \end{bmatrix}}_{\mathbf{V}^{K m_i m_j \times 2K m_i m_j}} = \mathbf{V}^\top \mathbf{V} \quad (6d)$$

The first equality repeats the definition of \mathcal{H}' in (3). The second equality uses the square root factorization of a symmetric matrix, which we define as

$$\mathbf{S}_{jj} = \mathbf{S}_{jj}^{\frac{1}{2}} \mathbf{S}_{jj}^{\frac{1}{2} \top} = \Theta^{\frac{1}{2}} \Theta^{\frac{1}{2} \top} = \Theta \Leftrightarrow \mathbf{S}_{jj}^{-1} = \mathbf{S}_{jj}^{-\frac{1}{2} \top} \mathbf{S}_{jj}^{-\frac{1}{2}} = \Theta^{-\frac{1}{2} \top} \Theta^{-\frac{1}{2}} = \Theta^{-1} \quad (7)$$

The third equality follows from Identity 5 on page 12, which leads directly to the desired factorization in the fourth step.

4.1.2 Find $\mathbf{x} \neq \mathbf{0}$ such that $\mathbf{V}\mathbf{x} = \mathbf{0}$

Next, we find a non-zero vector $\mathbf{x} \in \mathbb{R}^{2Km_i m_j \times 1}$ such that $\mathbf{V}\mathbf{x} = \mathbf{0}$. Without loss of generality, we look for \mathbf{x} in the general form

$$\mathbf{x} = \begin{bmatrix} \text{vec}\{\mathbf{M}^{[1]}\} \\ \vdots \\ \text{vec}\{\mathbf{M}^{[K]}\} \\ -\text{vec}\{\mathbf{N}^{[1]}\} \\ \vdots \\ -\text{vec}\{\mathbf{N}^{[K]}\} \end{bmatrix} = \begin{bmatrix} \text{vecbd}_\zeta\{\mathbf{M}\} \\ -\text{vecbd}_\zeta\{\mathbf{N}\} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu} \\ -\boldsymbol{\nu} \end{bmatrix} \quad (8)$$

where

$$\boldsymbol{\mu} = \text{vecbd}_\zeta\{\mathbf{M}\} = \begin{bmatrix} \text{vec}\{\mathbf{M}^{[1]}\} \\ \vdots \\ \text{vec}\{\mathbf{M}^{[K]}\} \end{bmatrix}, \quad \mathbf{M} \triangleq \bigoplus_{k=1}^K \mathbf{M}^{[k]} = \mathbf{M}, \quad \mathbf{M}^{[k]} \in \mathbb{R}^{m_i \times m_j},$$

$$\boldsymbol{\nu} = \text{vecbd}_\zeta\{\mathbf{N}\} = \begin{bmatrix} \text{vec}\{\mathbf{N}^{[1]}\} \\ \vdots \\ \text{vec}\{\mathbf{N}^{[K]}\} \end{bmatrix}, \quad \mathbf{N} \triangleq \bigoplus_{k=1}^K \mathbf{N}^{[k]} = \mathbf{N}, \quad \mathbf{N}^{[k]} \in \mathbb{R}^{m_i \times m_j}$$

The ‘‘vecbd’’ operator is defined in Definition A.2 on page 10. ζ stands for the block-partition $m_i \mathbf{1}_K \times m_j \mathbf{1}_K$. Setting $\mathbf{V}\mathbf{x} = \mathbf{0}$ implies that

$$\begin{bmatrix} \mathbf{S}_{jj}^{\frac{1}{2}\top} \boxminus \mathbf{S}_{ii}^{-\frac{1}{2}} & \mathbf{S}_{jj}^{-\frac{1}{2}} \boxminus \mathbf{S}_{ii}^{\frac{1}{2}\top} \end{bmatrix} \begin{bmatrix} \text{vecbd}_\zeta\{\mathbf{M}\} \\ -\text{vecbd}_\zeta\{\mathbf{N}\} \end{bmatrix} = \mathbf{0} \quad (9)$$

for some non-zero $\text{vecbd}_\zeta\{\mathbf{M}\}$ and $\text{vecbd}_\zeta\{\mathbf{N}\}$. Equality (9) can be rewritten as

$$(\mathbf{S}_{jj}^{\frac{1}{2}\top} \boxminus \mathbf{S}_{ii}^{-\frac{1}{2}}) \text{vecbd}_\zeta\{\mathbf{M}\} = (\mathbf{S}_{jj}^{-\frac{1}{2}} \boxminus \mathbf{S}_{ii}^{\frac{1}{2}\top}) \text{vecbd}_\zeta\{\mathbf{N}\} \quad (10)$$

We now turn to finding these $\text{vecbd}_\zeta\{\mathbf{M}\}$ and $\text{vecbd}_\zeta\{\mathbf{N}\}$ (alternatively: $\{\mathbf{M}^{[k]}\}_{k=1}^K$ and $\{\mathbf{N}^{[k]}\}_{k=1}^K$). Using Identity 3 on page 12, equality (10) can be rewritten as

$$\text{vec}\{\mathbf{S}_{ii}^{-\frac{1}{2}} \mathbf{M} \mathbf{S}_{jj}^{\frac{1}{2}}\} = \text{vec}\{\mathbf{S}_{ii}^{\frac{1}{2}\top} \mathbf{N} \mathbf{S}_{jj}^{-\frac{1}{2}\top}\}. \quad (11)$$

Removing the ‘‘vec’’ notation,

$$\mathbf{S}_{ii}^{-\frac{1}{2}} \mathbf{M} \mathbf{S}_{jj}^{\frac{1}{2}} = \mathbf{S}_{ii}^{\frac{1}{2}\top} \mathbf{N} \mathbf{S}_{jj}^{-\frac{1}{2}\top}. \quad (12)$$

Since \mathbf{S}_{ii} and \mathbf{S}_{jj} are invertible, the latter is equivalent to

$$\boxed{\mathbf{M}^{[k]} \mathbf{S}_{jj}^{[k,l]} = \mathbf{S}_{ii}^{[k,l]} \mathbf{N}^{[l]} \in \mathbb{R}^{m_i \times m_j} \quad \forall k, l} \quad (13a)$$

$$\boxed{\mathbf{M} \mathbf{S}_{jj} = \mathbf{S}_{ii} \mathbf{N}, \quad \mathbf{M} \triangleq \bigoplus_{k=1}^K \mathbf{M}^{[k]}, \quad \mathbf{N} \triangleq \bigoplus_{k=1}^K \mathbf{N}^{[k]}} \quad (13b)$$

Hence, our goal is to find non-zero $\{\mathbf{M}^{[k]}\}_{k=1}^K$ and/or $\{\mathbf{N}^{[k]}\}_{k=1}^K$ for which equality (13) holds.

4.1.3 From $\mathbf{M} \mathbf{S}_{jj} = \mathbf{S}_{ii} \mathbf{N}$ to $\mathbf{L} \mathbf{R} = \mathbf{P} \mathbf{L}$

The identifiability problem (13) can be further simplified into characterizing all the solutions to

$$\boxed{\mathbf{L}^{[k]} \mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]} \mathbf{L}^{[l]} \in \mathbb{R}^{P \times R} \quad \forall k, l} \quad (14a)$$

$$\boxed{\mathbf{L} \mathbf{R} = \mathbf{P} \mathbf{L}, \quad \mathbf{L} \triangleq \bigoplus_{k=1}^K \mathbf{L}^{[k]}} \quad (14b)$$

where $\mathbf{P} \in \mathbb{R}^{KP \times KP}$ and $\mathbf{R} \in \mathbb{R}^{KR \times KR}$ are normalized versions of \mathbf{S}_{ii} and \mathbf{S}_{jj} such that their (k, k) th main-diagonal blocks are equal to the identity:

$$\mathbf{P}^{[k,k]} = \mathbf{I}_P \quad \text{and} \quad \mathbf{R}^{[k,k]} = \mathbf{I}_R \quad (15)$$

and $P = m_i$, $R = m_j$. The normalization scheme that leads to (15) is explained in Appendix C. The positive definite matrices \mathbf{P} and \mathbf{R} are partitioned similarly to \mathbf{S}_{ii} and \mathbf{S}_{jj} into $K \times K$ blocks such that $\mathbf{P}^{[k,l]} \in \mathbb{R}^{P \times P}$ and $\mathbf{R}^{[k,l]} \in \mathbb{R}^{R \times R}$. Accordingly, $\mathbf{L}^{[k]} \in \mathbb{R}^{P \times R}$, $\mathbf{L} \triangleq \text{bdiag}\{\mathbf{L}^{[1]}, \dots, \mathbf{L}^{[K]}\} = \bigoplus_{k=1}^K \mathbf{L}^{[k]}$. Problem (14) is simpler than (13) since \mathbf{L} replaces both \mathbf{M} and \mathbf{N} , thus cutting by half the number of unknowns. The problem can now be reformulated as finding the minimal conditions on \mathbf{P} and \mathbf{R} such that $\{\mathbf{L}^{[k]}\}_{k=1}^K$ is non-zero and (14) holds, without breaking the admissibility constraints in Section 2.1.

Proof of (14). Applying the normalization scheme (63) to (13) leads to

$$\underbrace{\Omega_{ii}^{[k]} \mathbf{M}^{[k]} \Omega_{jj}^{-[k]}}_{\mathbf{L}^{[k]}} \mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]} \underbrace{\Omega_{ii}^{-[l]\top} \mathbf{N}^{[l]} \Omega_{jj}^{[l]\top}}_{\mathbf{L}'^{[k]}} \Leftrightarrow \underbrace{\Omega_{ii} \mathbf{M} \Omega_{jj}^{-1}}_{\mathbf{L}} \mathbf{R} = \mathbf{P} \underbrace{\Omega_{ii}^{-\dagger} \mathbf{N} \Omega_{jj}^{\top}}_{\mathbf{L}'} \quad (16)$$

The key point is that due to the normalization, for $l = k$,

$$\underbrace{\Omega_{ii}^{[k]} \mathbf{M}^{[k]} \Omega_{jj}^{-[k]}}_{\mathbf{L}^{[k]}} \underbrace{\mathbf{R}^{[k,k]}}_{\mathbf{I}} = \underbrace{\mathbf{P}^{[k,k]}}_{\mathbf{I}} \underbrace{\Omega_{ii}^{-[k]\top} \mathbf{N}^{[k]} \Omega_{jj}^{[k]\top}}_{\mathbf{L}'^{[k]}} \quad (17)$$

Equation (17) implies that $\mathbf{L}^{[k]} = \mathbf{L}'^{[k]} \forall k$, which concludes the proof. \square

4.2 Main Result: JISA Identifiability

It is clear that (14) is identical to (47), as well as to the main equation in Lemma 2–3, up to the arbitrary normalization. Therefore, the admissible solutions to (14) are given by Lemma 2–3. Non-identifiability of admissible data (Definition 2.1) is associated with the following scenarios. Their interpretation in terms of the unnormalized sources is given in Theorem 4.1.

Scenario 1. The first type of non-identifiability corresponds to Scenario 6 on page 16. Given a pair of sources with the following structure of their covariance matrices,

$$\mathbf{S}_{ii} = \left[\begin{array}{c|c} \mathbf{S}_{ii}^{[1:D,1:D]} & \mathbf{0}_{Dm_i \times (K-D)m_i} \\ \hline \mathbf{0}_{(K-D)m_i \times Dm_i} & \mathbf{S}_{ii}^{[D+1:K,D+1:K]} \end{array} \right] \quad \text{and} \quad \mathbf{S}_{jj} = \left[\begin{array}{c|c} \mathbf{S}_{jj}^{[1:D,1]} & \mathbf{0}_{Dm_j \times (K-D)m_j} \\ \hline \mathbf{0}_{(K-D)m_j \times Dm_j} & \mathbf{S}_{jj}^{[D+1:K,D+1:K]} \end{array} \right]$$

non-identifiability exists if and only if $m_i = m_j$ and the sources are linked by

$$\mathbf{S}_{jj}^{[1:D,1:D]} = \text{bdiag}\{\Psi^{[1]}, \dots, \Psi^{[D]}\} \mathbf{S}_{ii}^{[1:D,1:D]} \text{bdiag}^{\top}\{\Psi^{[1]}, \dots, \Psi^{[D]}\} \quad (18)$$

for any D invertible $m_i \times m_i$ matrices $\Psi^{[k]}$, $k = 1, \dots, D$. Equation (18) can be regarded as a generalized similarity transformation, or an equivalence relation, in the sense of Definition 2.3 on page 3, between $\mathbf{S}_{ii}^{[1:D,1:D]}$ and $\mathbf{S}_{jj}^{[1:D,1:D]}$.

Scenario 2. The second type of non-identifiability is associated with Scenario 7 on page 16. In this case, the model is non-identifiable if there exists a pair of sources of possibly different dimensions (without loss of generality, $m_i \geq m_j$) with structure

$$\mathbf{S}_{ii} = \left[\begin{array}{ccc|c} \mathbf{S}_{ii}^{[1,1]} & \dots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \mathbf{0}_{Dm_i \times (K-D)m_i} \\ \mathbf{0} & \dots & \mathbf{S}_{ii}^{[D,D]} & \\ \hline & \mathbf{0}_{(K-D)m_i \times Dm_i} & & \mathbf{S}_{ii}^{[D+1:K,D+1:K]} \end{array} \right] \quad \text{and} \quad \mathbf{S}_{jj} = \left[\begin{array}{ccc|c} \mathbf{S}_{jj}^{[1,1]} & \dots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \mathbf{0}_{Dm_j \times (K-D)m_j} \\ \mathbf{0} & \dots & \mathbf{S}_{jj}^{[D,D]} & \\ \hline & \mathbf{0}_{(K-D)m_j \times Dm_j} & & \mathbf{S}_{jj}^{[D+1:K,D+1:K]} \end{array} \right]$$

Theorem 4.1 (JISA non-identifiability). *The JISA model is not identifiable if and only if (iff) there exists at least one pair (i, j) of positive definite covariance matrices \mathbf{S}_{jj} and \mathbf{S}_{ii} of dimensions $Km_j \times Km_j$ and $Km_i \times Km_i$, respectively, admissible by Definition 2.1, for which either*

1. *there exists an index k' such that $\forall k \neq k'$, $\mathbf{S}_{ii}^{[k',k]} = \mathbf{0}_{m_i \times m_i} = \mathbf{S}_{ii}^{[k,k']}$ and $\mathbf{S}_{jj}^{[k',k]} = \mathbf{0}_{m_j \times m_j} = \mathbf{S}_{jj}^{[k,k']}$ (that is, a pair of sources does not have correlated counterparts in any of the other mixtures),*
or
2. (a) $m_i = m_j$ and

(b) there exists a number $K' > 1$ and an ordering of the mixtures such that

$$\mathbf{S}_{ii}^{[1:K', k > K']} = \mathbf{0}_{m_i \times m_i} = \mathbf{S}_{ii}^{[k > K', 1:K']} = \mathbf{S}_{jj}^{[1:K', k > K']} = \mathbf{S}_{jj}^{[k > K', 1:K']}$$

(the above is equivalent to \mathbf{S}_{ii} and \mathbf{S}_{jj} being block-diagonal with $\mathbf{S}_{ii}^{[1:K', 1:K']}$ and $\mathbf{S}_{jj}^{[1:K', 1:K]}$ on their first diagonal block, respectively, under some permutation of the mixture indices) and

$$\mathbf{S}_{jj}^{[1:K', 1:K']} = \text{bdiag}\{\Psi^{[1]}, \dots, \Psi^{[K']}\} \mathbf{S}_{ii}^{[1:K', 1:K']} \text{bdiag}^\top\{\Psi^{[1]}, \dots, \Psi^{[K']}\}$$

where $\Psi^{[k]}$ are arbitrary invertible $m_i \times m_i$ matrices.

5 Discussion

In this paper, we presented new results in algebra and in signal processing. This was achieved by analysing the FIM of a recently-proposed source separation model that is inspired by data fusion. Therefore, one of our messages in this work is that by formulating new methods in which data sets can interact, we obtain new types of algebraic structures, and the theoretical analysis of these algebraic structures yields new insights and contributions that go beyond their community of origin.

This algebraic result, formulated in several lemmas, can be regarded as a generalization to and a variation of Schur's lemma on irreducible representations. This lemma was used to derive the nongeneric uniqueness and identifiability conditions of JISA, when SOS are involved and the mixing is invertible. This model, as well as the corresponding lemmas, can be extended by further relaxing some of the numerical and structural assumptions.

From a data fusion perspective, the JISA model is non-identifiable in two main scenarios: first, if there exists at least one pair of sources with equivalent subspaces, in the generalized sense. Second, if there exists a pair of sources with no counterparts in the other data sets. All other scenarios are identifiable, if the data is admissible. This implies that JISA can be used for data fusion even if there are only very few links among corresponding sources in different data sets.

Finally, Table 1 provides further insights into our new results by comparing the original and new lemmas. "Variation" in the first row implies that we use symmetric matrices instead of irreducible representations of symmetry groups, as in the original formulation. Table 1 clarifies why we call the new formulation "generalized": when all commuting matrices are forced to be identical, we obtain the original Schur formulation.

		Multiset analogue (new)	Schur's lemma (variation)
First lemma	Input data	$k, l = 1, \dots, K$ $\mathbf{M}^{[k]}, \mathbf{C}^{[k,l]} P \times P$ $\mathbf{C}^{[k,l]} = (\mathbf{C}^{[l,k]})^\top$	$q = 1, \dots, Q$ $\mathbf{M}, \mathbf{C}^{(q)} P \times P$ $\mathbf{C}^{(q)} = (\mathbf{C}^{(q)})^\top$
	Commutation Non-uniqueness Comm. if $\mathbf{M}^{[k]} = \mathbf{M} \forall k$	$\mathbf{M}^{[k]} \mathbf{C}^{[k,l]} = \mathbf{C}^{[k,l]} \mathbf{M}^{[l]} \forall k, l$ $\mathbf{M}^{[k]} = \mu \mathbf{I}_P \forall k, \mu \geq 0$ $\mathbf{M} \mathbf{C}^{[k,l]} = \mathbf{C}^{[k,l]} \mathbf{M} \forall k, l$	$\mathbf{M} \mathbf{C}^{(q)} = \mathbf{C}^{(q)} \mathbf{M} \forall q$ $\mathbf{M} = \mu \mathbf{I}_P, \mu \geq 0$
Second lemma	Input data	$\mathbf{M}^{[k]} P' \times P'$ $\mathbf{C}^{[k,l]} P \times P, \mathbf{C}'^{[k,l]} P' \times P'$ $P \times P, \mathbf{C}'^{(q)} P' \times P'$	$\mathbf{M} P' \times P'$ $\mathbf{C}^{(q)}$
	Commutation Non-uniqueness Comm. if $\mathbf{M}^{[k]} = \mathbf{M} \forall k$	$\mathbf{M}^{[k]} \mathbf{C}^{[k,l]} = \mathbf{C}'^{[k,l]} \mathbf{M}^{[l]} \forall k, l$ $\mathbf{M}^{[k]} = \mathbf{0}_{P' \times P}$ or $\mathbf{M}^{[k]} = \lambda \mathbf{O}^{[k]} \forall k$ $\mathbf{M} \mathbf{C}^{[k,l]} = \mathbf{C}'^{[k,l]} \mathbf{M} \forall k, l$	$\mathbf{M} \mathbf{C}^{(q)} = \mathbf{C}'^{(q)} \mathbf{M} \forall q$ $\mathbf{M} = \mathbf{0}_{P' \times P}$ or $\mathbf{M} = \lambda \mathbf{O}$

Table 1: Comparison of Schur's lemma and its proposed multiset analogue, strict sense

A Some Algebraic Properties

For ease of reference, we list some useful algebraic properties. Properties that are not proved below can be found in [27, 28, 26]. A glossary of notations is given in Table 2.

Product Name	Notation	LaTeX Command
Hadamard	\otimes	<code>\had</code>
Khatri-Rao columnwise	\odot	<code>\khat</code>
Khatri-Rao block-columnwise	\boxtimes	<code>\khatcb</code>
Khatri-Rao for partitioned matrices / block-Kronecker	\boxplus	<code>\khatb</code>

Table 2: Glossary

For any matrices $\mathbf{M}, \mathbf{N}, \mathbf{P}, \mathbf{Q}$ (with appropriate dimensions),

$$(\mathbf{N} \otimes \mathbf{M})(\mathbf{P} \otimes \mathbf{Q}) = \mathbf{NP} \otimes \mathbf{MQ} \quad (19a)$$

$$(\mathbf{N} \otimes \mathbf{M})^\top = \mathbf{N}^\top \otimes \mathbf{M}^\top \quad (19b)$$

$$\text{vec}\{\mathbf{MQN}\} = (\mathbf{N}^\top \otimes \mathbf{M})\text{vec}\{\mathbf{Q}\} \quad (19c)$$

$$\text{tr}\{\mathbf{PQ}\} = \text{tr}\{\mathbf{QP}\} \quad (19d)$$

$$\text{tr}\{\mathbf{P}^\top \mathbf{Q}\} = \text{vec}^\dagger\{\mathbf{P}\}\text{vec}\{\mathbf{Q}\} \quad (19e)$$

$$\det(\mathbf{MN}) = \det(\mathbf{NM}). \quad (19f)$$

For any two matrices $\mathbf{M}_{M \times P}$ and $\mathbf{N}_{N \times Q}$,

$$\mathcal{T}_{M,N}(\mathbf{N} \otimes \mathbf{M}) = (\mathbf{M} \otimes \mathbf{N})\mathcal{T}_{P,Q}. \quad (20a)$$

Identity 1.

$$(\mathbf{I} \otimes \mathcal{T}_{m_i, m_j})(\mathbf{A} \boxplus \mathbf{B})(\mathbf{I} \otimes \mathcal{T}_{m_i, m_j}^\top) = \mathbf{B} \boxplus \mathbf{A} \quad (21)$$

Proof of Identity 1.

$$\begin{aligned}
& (\mathbf{I} \otimes \mathcal{T}_{m_i, m_j})(\mathbf{A} \boxplus \mathbf{B})(\mathbf{I} \otimes \mathcal{T}_{m_i, m_j}^\top) \\
&= \begin{bmatrix} \mathcal{T}_{m_i, m_j} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathcal{T}_{m_i, m_j} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} \otimes \mathbf{B}_{11} & \cdots & \mathbf{A}_{1K} \otimes \mathbf{B}_{1K} \\ \vdots & & \vdots \\ \mathbf{A}_{K1} \otimes \mathbf{B}_{K1} & \cdots & \mathbf{A}_{KK} \otimes \mathbf{B}_{KK} \end{bmatrix} \begin{bmatrix} \mathcal{T}_{m_i, m_j}^\top & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathcal{T}_{m_i, m_j}^\top \end{bmatrix} \\
&= \begin{bmatrix} \mathcal{T}_{m_i, m_j}(\mathbf{A}_{11} \otimes \mathbf{B}_{11})\mathcal{T}_{m_i, m_j}^\top & \cdots & \mathcal{T}_{m_i, m_j}(\mathbf{A}_{1K} \otimes \mathbf{B}_{1K})\mathcal{T}_{m_i, m_j}^\top \\ \vdots & & \vdots \\ \mathcal{T}_{m_i, m_j}(\mathbf{A}_{K1} \otimes \mathbf{B}_{K1})\mathcal{T}_{m_i, m_j}^\top & \cdots & \mathcal{T}_{m_i, m_j}(\mathbf{A}_{KK} \otimes \mathbf{B}_{KK})\mathcal{T}_{m_i, m_j}^\top \end{bmatrix} \\
&\stackrel{(20a)}{=} \begin{bmatrix} \mathbf{B}_{11} \otimes \mathbf{A}_{11} & \cdots & \mathbf{B}_{1K} \otimes \mathbf{A}_{1K} \\ \vdots & & \vdots \\ \mathbf{B}_{K1} \otimes \mathbf{A}_{K1} & \cdots & \mathbf{B}_{KK} \otimes \mathbf{A}_{KK} \end{bmatrix} = \mathbf{B} \boxplus \mathbf{A} \quad (22)
\end{aligned}$$

□

A.1 diag, bdiag, vecd, vecbd operators

In order to avoid confusion with the vecd and vecbd operators, we define

$$\begin{aligned}
\text{diag}\{\mathbf{X}\} &\triangleq \begin{bmatrix} x_{11} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_{KK} \end{bmatrix} = \text{diag}(\text{diag}(\mathbf{X})), \quad \mathbf{X} \in \mathbb{R}^{K \times K} \\
\text{diag}\{\mathbf{x}\} &\triangleq \begin{bmatrix} x_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & x_K \end{bmatrix} = \text{diag}(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^{K \times 1}
\end{aligned}$$

Now, let $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^\top$, $\sum_{k=1}^K \alpha_k = \alpha$ and similarly for β . Then,

$$\begin{aligned}
\text{bdiag}_{\boldsymbol{\alpha} \times \boldsymbol{\beta}}\{\mathbf{X}\} &\triangleq \begin{bmatrix} \mathbf{X}_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X}_{KK} \end{bmatrix} = \bigoplus_{k=1}^K \mathbf{X}_{kk} = \text{bdiag}\{\mathbf{X}_{11}, \dots, \mathbf{X}_{KK}\}, \quad \mathbf{X}_{kk} \in \mathbb{R}^{\alpha_k \times \beta_k} \\
\text{bdiag}_{\boldsymbol{\alpha}}\{\mathbf{X}\} &\triangleq \text{bdiag}_{\boldsymbol{\alpha} \times \boldsymbol{\alpha}}\{\mathbf{X}\}, \quad \mathbf{X} \in \mathbb{R}^{\alpha \times \alpha}
\end{aligned}$$

Definition A.1 (vecd Operator). For any square matrix $\mathbf{X} \in \mathbb{R}^{K \times K}$ with entries x_{ij} , $i, j \in 1, \dots, K$, define the operator

$$\text{vecd}\{\mathbf{X}\} \triangleq \begin{bmatrix} x_{11} \\ \vdots \\ x_{KK} \end{bmatrix} = \text{diag}(\mathbf{X}) \in \mathbb{R}^{K \times 1}. \quad (23)$$

That is, $\text{vecd}\{\mathbf{X}\}$ is a vector that consists only of the entries on the diagonal of \mathbf{X} .

Definition A.2 (vecbd Operator). For any rectangular matrix $\mathbf{X} \in \mathbb{R}^{\alpha \times \beta}$ partitioned into K rows and K columns such that its (i, j) th block is $\mathbf{X}_{ij} \in \mathbb{R}^{\alpha_k \times \beta_k}$, $i, j \in 1, \dots, K$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^\top$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]^\top$, $\alpha = \sum_{k=1}^K \alpha_k$, $\beta = \sum_{k=1}^K \beta_k$, define the operator

$$\text{vecbd}_{\boldsymbol{\alpha} \times \boldsymbol{\beta}}\{\mathbf{X}\} \triangleq \begin{bmatrix} \text{vec}\{\mathbf{X}_{11}\} \\ \vdots \\ \text{vec}\{\mathbf{X}_{KK}\} \end{bmatrix} \in \mathbb{R}^{(\sum_{k=1}^K \alpha_k \beta_k) \times 1} \neq \text{vec}\{\text{bdiag}_{\boldsymbol{\alpha} \times \boldsymbol{\beta}}\{\mathbf{X}\}\} \quad (24)$$

That is, $\text{vecbd}_{\boldsymbol{\alpha} \times \boldsymbol{\beta}}\{\mathbf{X}\}$ is a vector that consists only of the (vectorized) entries of the block-diagonal of \mathbf{X} , where the rows of \mathbf{X} are partitioned according to $\boldsymbol{\alpha}$ and the columns by $\boldsymbol{\beta}$. If $\boldsymbol{\alpha} = \boldsymbol{\beta}$ then we can write

$$\text{vecbd}_{\boldsymbol{\alpha}}\{\mathbf{X}\} \triangleq \text{vecbd}_{\boldsymbol{\alpha} \times \boldsymbol{\alpha}}\{\mathbf{X}\}. \quad (25)$$

A.2 Khatri-Rao, Kronecker, Tensor Matricization and Vectorization

Tensors can be written as multidimensional arrays, matricized or vectorized. Even within these representations, there are variations. Consider a third-order tensor whose representation in multilinear products is

$$\mathcal{T} = \mathcal{D} \times_1 \mathbf{A}' \times_2 \mathbf{B}' \times_3 \mathbf{C}' \quad (\text{CPD}) \quad (26a)$$

$$\mathcal{T} = \mathcal{G} \times_1 \mathbf{A} \times_2 \mathbf{B} \times_3 \mathbf{C} \quad (\text{Tucker}) \quad (26b)$$

where \mathcal{D} is a tensor with diagonal core \mathbf{d} [29, Table I, Eq. (4)]. Then, its vector representation can be written as [29, Table III]

$$\text{vec}\{\mathcal{T}\} = (\mathbf{C}' \odot \mathbf{B}' \odot \mathbf{A}')\mathbf{d} \quad (\text{CPD}) \quad (27a)$$

$$\text{vec}\{\mathcal{T}\} = (\mathbf{C} \otimes \mathbf{B} \otimes \mathbf{A})\text{vec}\{\mathcal{G}\} \quad (\text{Tucker}) \quad (27b)$$

If now \mathcal{T} is a second-order tensor, then (26) reduces to

$$\mathbf{T} = \mathbf{D} \times_1 \mathbf{A}' \times_2 \mathbf{B}' = \mathbf{A}'\mathbf{D}\mathbf{B}'^\top, \quad \mathbf{D} = \text{diag}\{\mathbf{d}\} \quad (28a)$$

$$\mathbf{T} = \mathbf{G} \times_1 \mathbf{A} \times_2 \mathbf{B} = \mathbf{A}\mathbf{G}\mathbf{B}^\top \quad (28b)$$

and (27) to

$$\text{vec}\{\mathbf{T}\} = (\mathbf{B}' \odot \mathbf{A}')\mathbf{d} \quad (29a)$$

$$\text{vec}\{\mathbf{T}\} = (\mathbf{B} \otimes \mathbf{A})\text{vec}\{\mathbf{G}\} \quad (29b)$$

Combining (28) with (29), and using the notation $\mathbf{d} = \text{vecd}\{\mathbf{D}\}$ from Definition A.1, we obtain

$$\text{vec}\{\mathbf{T}\} = \overbrace{(\mathbf{B}' \odot \mathbf{A}')\text{vecd}\{\mathbf{D}\}}^{\text{Only for } \mathbf{D} \in \text{diag}} = \text{vec}\{\mathbf{A}'\mathbf{D}\mathbf{B}'^\top\} \quad (30a)$$

$$\text{vec}\{\mathbf{T}\} = \underbrace{(\mathbf{B} \otimes \mathbf{A})\text{vec}\{\mathbf{G}\}}_{\text{Well-known identity } \forall \mathbf{G}} = \text{vec}\{\mathbf{A}\mathbf{G}\mathbf{B}^\top\} \quad (30b)$$

The equalities in (30) are summarized in the following identity

Identity 2. For any $\mathbf{X} \in \text{diag}$, and matrices \mathbf{A} and \mathbf{B} with appropriate dimensions,

$$(\mathbf{B} \odot \mathbf{A})\text{vecd}\{\mathbf{X}\} \stackrel{\mathbf{X} \in \text{diag}}{=} (\mathbf{B} \otimes \mathbf{A})\text{vec}\{\mathbf{X}\} \stackrel{\forall \mathbf{X}}{=} \text{vec}\{\mathbf{A}\mathbf{X}\mathbf{B}^\top\} \quad (31)$$

where the operator “ $\text{vecd}\{\cdot\}$ ” was defined in Definition A.1. The second equality is true for any \mathbf{X} , and the first one only for \mathbf{X} diagonal.

Identity 2 appears in Brewer [30, Table III, T3.13], as well as in Liu and Trenkler [31, Equation (27)].

Similarly to the argumentations in (26)–(30), we can also vectorize a Tucker format of a second-order tensor when the core tensor is block-diagonal.

Identity 3. Let $\mathbf{A} \in \mathbb{R}^{\mu \times \alpha}$ and $\mathbf{B} \in \mathbb{R}^{\nu \times \beta}$ be two matrices partitioned into K column blocks of dimensions $\mu \times \alpha_k$ and $\nu \times \beta_k$, respectively, $\alpha = \sum_{k=1}^K \alpha_k$, $\beta = \sum_{k=1}^K \beta_k$, $\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_K]^\top$, $\boldsymbol{\beta} = [\beta_1, \dots, \beta_K]^\top$, as follows,

$$\begin{aligned} \mathbf{A} &= [\mathbf{A}_1 \mid \cdots \mid \mathbf{A}_K], & \mathbf{A}_k &\in \mathbb{R}^{\mu \times \alpha_k} \\ \mathbf{B} &= [\mathbf{B}_1 \mid \cdots \mid \mathbf{B}_K], & \mathbf{B}_k &\in \mathbb{R}^{\nu \times \beta_k} \end{aligned} \quad (32)$$

and $\mathbf{X} = \bigoplus_{k=1}^K \mathbf{X}_{kk} \in \mathbb{R}^{\alpha \times \beta}$, $\mathbf{X}_{kk} \in \mathbb{R}^{\alpha_k \times \beta_k}$. Then

$$(\mathbf{B} \square \mathbf{A}) \text{vecbd}_{\boldsymbol{\alpha} \times \boldsymbol{\beta}} \{ \mathbf{X} \} = \text{vec} \{ \mathbf{A} \mathbf{X} \mathbf{B}^\top \}, \quad (33)$$

where the operator “ $\text{vecbd}_{\boldsymbol{\alpha} \times \boldsymbol{\beta}} \{ \cdot \}$ ” was defined in Definition A.2 and “ \square ” in Table 2.

Remark 1. Identity 3 is a generalization of Identity 2.

Proof of Identity 3 and Equation (33). Apart from the Tucker format vectorization, which is a constructive proof, we can also prove directly. The following proof is based on the fact that

$$\mathbf{A}_k = \mathbf{A} \mathbf{J}_{\alpha_k} \quad (34)$$

where $\mathbf{J}_{\alpha_k} = \begin{bmatrix} \mathbf{0} \\ \mathbf{I}_{\alpha_k} \\ \mathbf{0} \end{bmatrix} \in \mathbb{R}^{\alpha \times \alpha_k}$ is a matrix of zeros, with \mathbf{I}_{α_k} at the rows required to extract the columns pertaining to \mathbf{A}_k from \mathbf{A} . We define a similar matrix for \mathbf{B} . Then,

$$\begin{aligned} (\mathbf{B} \square \mathbf{A}) \text{vecbd}_{\boldsymbol{\alpha} \times \boldsymbol{\beta}} \{ \mathbf{X} \} &= [\mathbf{B} \mathbf{J}_{\beta_1} \otimes \mathbf{A} \mathbf{J}_{\alpha_1} \mid \cdots \mid \mathbf{B} \mathbf{J}_{\beta_K} \otimes \mathbf{A} \mathbf{J}_{\alpha_K}] \begin{bmatrix} \text{vec} \{ \mathbf{X}_{11} \} \\ \vdots \\ \text{vec} \{ \mathbf{X}_{KK} \} \end{bmatrix} \\ &= \sum_{k=1}^K (\mathbf{B}^\top \mathbf{J}_{\beta_k} \otimes \mathbf{A} \mathbf{J}_{\alpha_k}) \text{vec} \{ \mathbf{X}_{kk} \} \stackrel{(19c)}{=} \sum_{k=1}^K \text{vec} \{ \mathbf{A} \mathbf{J}_{\alpha_k} \mathbf{X}_{kk} \mathbf{J}_{\beta_k}^\top \mathbf{B} \} \\ &= \text{vec} \{ \mathbf{A} \underbrace{ \left(\sum_{k=1}^K \mathbf{J}_{\alpha_k} \mathbf{X}_{kk} \mathbf{J}_{\beta_k}^\top \right) }_{\bigoplus_{k=1}^K \mathbf{X}_{kk} = \mathbf{X}} \mathbf{B}^\top \} = \text{vec} \{ \mathbf{A} \mathbf{X} \mathbf{B}^\top \} \end{aligned} \quad (35)$$

□

Identity 4. Consider four matrices $\mathbf{A} \in \mathbb{R}^{\mu \times \xi}$, $\mathbf{B} \in \mathbb{R}^{\nu \times \zeta}$, $\mathbf{C} \in \mathbb{R}^{\alpha \times \eta}$, $\mathbf{D} \in \mathbb{R}^{\gamma \times \delta}$. Then,

$$\underbrace{\underbrace{\underbrace{\mathbf{A}^\top \mathbf{C}}_{\Rightarrow \mu = \alpha}}_{\Rightarrow \xi = \zeta, \eta = \delta} \circledast \underbrace{\underbrace{\mathbf{B}^\top \mathbf{D}}_{\Rightarrow \nu = \gamma}}_{\Rightarrow \eta = \delta}}_{\Rightarrow \xi = \zeta, \eta = \delta} = \underbrace{\underbrace{\underbrace{\mathbf{A} \circledast \mathbf{B}}_{\Rightarrow \xi = \zeta}}_{\xi \times \mu \nu}}_{\Rightarrow \mu \nu = \alpha \gamma} \underbrace{\underbrace{\mathbf{C} \circledast \mathbf{D}}_{\Rightarrow \eta = \delta}}_{\alpha \gamma \times \eta} \quad (36)$$

where “ \circledast ” denotes the “columnwise Khatri-Rao product” and “ \circledast ” the scalar Hadamard product, see Table 2. We see that the constraints $\xi = \zeta$ and $\eta = \delta$ occur on both sides of (36). The constraint on the right-hand side (RHS) “ $\mu \nu = \alpha \gamma$ ” is fulfilled with the left-hand side (LHS) constraints $\mu = \alpha$ and $\nu = \gamma$. We thus remain with

$$\mathbf{A}^\top \mathbf{C} \circledast \mathbf{B}^\top \mathbf{D} = (\mathbf{A}_{\mu \times \xi} \circledast \mathbf{B}_{\nu \times \zeta})^\top (\mathbf{C}_{\mu \times \eta} \circledast \mathbf{D}_{\nu \times \delta}). \quad (37)$$

The constraints on the dimensions of the matrices imply that they may be regarded as sub-blocks of the same matrix,

$$\begin{bmatrix} \mathbf{A}_{\mu \times \xi} & \mathbf{C}_{\mu \times \eta} \\ \mathbf{B}_{\nu \times \zeta} & \mathbf{D}_{\nu \times \delta} \end{bmatrix} \quad (38)$$

Proof of Identity 4 and (36). On the LHS of (36), the (i, j) th scalar entry of $\mathbf{A}^\top \mathbf{C}$ is

$$[\mathbf{A}^\top \mathbf{C}]_{ij} = \sum_{k=1}^{\mu} a_{ki} c_{kj} = (\mathbf{a}_i)^\top \mathbf{c}_j \quad (39)$$

where a_{ki} is the (k, i) th scalar entry of \mathbf{A} and \mathbf{a}_i is the i th column vector of \mathbf{A} . Hence,

$$[\mathbf{A}^\top \mathbf{C} \otimes \mathbf{B}^\top \mathbf{D}]_{ij} = (\mathbf{a}_i)^\top \mathbf{c}_j \cdot (\mathbf{b}_i)^\top \mathbf{d}_j \quad (40)$$

On the RHS of (36),

$$(\mathbf{A} \odot \mathbf{B})^\top (\mathbf{C} \odot \mathbf{D}) = \begin{bmatrix} (\mathbf{a}_1 \otimes \mathbf{b}_1)^\top \\ \vdots \\ (\mathbf{a}_K \otimes \mathbf{b}_K)^\top \end{bmatrix} [\mathbf{c}_1 \otimes \mathbf{d}_1 \quad \cdots \quad \mathbf{c}_K \otimes \mathbf{d}_K] \quad (41)$$

The (i, j) th block of (41) is

$$(\mathbf{a}_i \otimes \mathbf{b}_i)^\top (\mathbf{c}_j \otimes \mathbf{d}_j) = (\mathbf{a}_i^\top \mathbf{c}_j) \otimes (\mathbf{b}_i^\top \mathbf{d}_j) = (\mathbf{a}_i^\top \mathbf{c}_j)(\mathbf{b}_i^\top \mathbf{d}_j) = (40) \quad (42)$$

□

Remark 2. Identity 4 is a special case of Identity 5.

Remark 3. Equation (36) is our extension to

$$\mathbf{A}^\top \mathbf{A} \otimes \mathbf{B}^\top \mathbf{B} = (\mathbf{A} \odot \mathbf{B})^\top (\mathbf{A} \odot \mathbf{B}), \quad (43)$$

see e.g. [31, 32, 33, 34].

In order to solve the multidimensional case, we need to extend Identity 4 (Equation (36)) and Identity 2 (Equation (31)) to any block-partitions.

Identity 5. Let $\mathbf{A} \in \mathbb{R}^{\mu \times a}$, $\mathbf{B} \in \mathbb{R}^{\phi \times b}$, $\mathbf{C} \in \mathbb{R}^{\mu \times c}$, $\mathbf{D} \in \mathbb{R}^{\nu \times d}$ be four matrices partitioned to column blocks as $\mathbf{A}_k \in \mathbb{R}^{\mu \times \alpha_k}$, $\mathbf{B}_k \in \mathbb{R}^{\phi \times \beta_k}$, $\mathbf{C} \in \mathbb{R}^{\nu \times \gamma_k}$, $\mathbf{D} \in \mathbb{R}^{\nu \times \delta_k}$, $\sum_{k=1}^K \alpha_k = a$, $\sum_{k=1}^K \beta_k = b$, $\sum_{k=1}^K \gamma_k = c$, $\sum_{k=1}^K \delta_k = d$. Then,

$$(\mathbf{A} \boxplus \mathbf{B})^\top (\mathbf{C} \boxplus \mathbf{D}) = \mathbf{A}^\top \mathbf{C} \boxplus \mathbf{B}^\top \mathbf{D} \quad (44)$$

Proof of Identity 5 and (44).

$$\begin{aligned} \mathbf{A}_{\mu \times a} \boxplus \mathbf{B}_{\nu \times b} &= [\mathbf{A}_1 \mid \cdots \mid \mathbf{A}_K] \boxplus [\mathbf{B}_1 \mid \cdots \mid \mathbf{B}_K] \\ &= [\mathbf{A}_1 \otimes \mathbf{B}_1 \mid \cdots \mid \underbrace{\mu\nu \times \alpha_k \beta_k}_{\mathbf{A}_k \otimes \mathbf{B}_k} \mid \cdots \mid \mathbf{A}_K \otimes \mathbf{B}_K] \in \mathbb{R}^{\mu\nu \times \sum_{k=1}^K \alpha_k \beta_k} \end{aligned} \quad (45)$$

Therefore,

$$\begin{aligned} (\mathbf{A} \boxplus \mathbf{B})^\top (\mathbf{C} \boxplus \mathbf{D}) &= \overbrace{[\mathbf{A}_1 \otimes \mathbf{B}_1 \mid \cdots \mid \mathbf{A}_K \otimes \mathbf{B}_K]^\top}^{\sum_{k=1}^K \alpha_k \beta_k \times \mu\nu} \overbrace{[\mathbf{C}_1 \otimes \mathbf{D}_1 \mid \cdots \mid \mathbf{C}_K \otimes \mathbf{D}_K]}^{\mu\nu \times \sum_{k=1}^K \gamma_k \delta_k} \\ &= \begin{bmatrix} (\mathbf{A}_1 \otimes \mathbf{B}_1)^\top \\ \vdots \\ (\mathbf{A}_K \otimes \mathbf{B}_K)^\top \end{bmatrix} [\mathbf{C}_1 \otimes \mathbf{D}_1 \mid \cdots \mid \mathbf{C}_K \otimes \mathbf{D}_K] \\ &= \begin{bmatrix} \mathbf{A}_1^\top \otimes \mathbf{B}_1^\top \\ \vdots \\ \mathbf{A}_K^\top \otimes \mathbf{B}_K^\top \end{bmatrix} [\mathbf{C}_1 \otimes \mathbf{D}_1 \mid \cdots \mid \mathbf{C}_K \otimes \mathbf{D}_K] \\ &= \begin{bmatrix} \mathbf{A}_1^\top \mathbf{C}_1 \otimes \mathbf{B}_1^\top \mathbf{D}_1 & \cdots & \mathbf{A}_1^\top \mathbf{C}_l \otimes \mathbf{B}_1^\top \mathbf{D}_l & \cdots & \mathbf{A}_1^\top \mathbf{C}_K \otimes \mathbf{B}_1^\top \mathbf{D}_K \\ \vdots & & & & \vdots \\ \mathbf{A}_k^\top \mathbf{C}_1 \otimes \mathbf{B}_k^\top \mathbf{D}_1 & \cdots & \underbrace{\mathbf{A}_k^\top \mathbf{C}_l \otimes \mathbf{B}_k^\top \mathbf{D}_l}_{\substack{[\mathbf{A}^\top \mathbf{C}]_{kl} \in \mathbb{R}^{\alpha_k \times \gamma_l} \quad [\mathbf{B}^\top \mathbf{D}]_{kl} \in \mathbb{R}^{\beta_k \times \delta_l} \\ \alpha_k \beta_k \times \gamma_l \delta_l}} & \cdots & \mathbf{A}_k^\top \mathbf{C}_K \otimes \mathbf{B}_k^\top \mathbf{D}_K \\ \vdots & & & & \vdots \\ \mathbf{A}_K^\top \mathbf{C}_1 \otimes \mathbf{B}_K^\top \mathbf{D}_1 & \cdots & \mathbf{A}_K^\top \mathbf{C}_l \otimes \mathbf{B}_K^\top \mathbf{D}_l & \cdots & \mathbf{A}_K^\top \mathbf{C}_K \otimes \mathbf{B}_K^\top \mathbf{D}_K \end{bmatrix} \end{aligned} \quad (46)$$

In the above, we only need that \mathbf{A} and \mathbf{C} and \mathbf{B} and \mathbf{D} have the same number of rows (this can be formulated as $[\mathbf{A} \ \mathbf{C}]$ and $[\mathbf{B} \ \mathbf{D}]$, less constrained than (38)) and that all matrices are partitioned into the same number K of block columns. We do not need that $\alpha_k, \beta_k, \gamma_k, \delta_k$ be equal. The (k, l) th block of (46) is exactly the (k, l) th block of $\mathbf{A}^\top \mathbf{C} \boxplus \mathbf{B}^\top \mathbf{D}$. □

B Proof of Lemma 2–3

Consider a set of matrices that commute as follows,

$$\mathbf{L}^{[k]} \mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]} \mathbf{L}^{[l]} \in \mathbb{R}^{P \times R} \quad \forall k, l \quad (47a)$$

$$\mathbf{L} \mathbf{R} = \mathbf{P} \mathbf{L} \quad , \quad \mathbf{L} \triangleq \bigoplus_{k=1}^K \mathbf{L}^{[k]} \quad (47b)$$

and assume also normalization of the form $\mathbf{P}^{[k,k]} = \mathbf{I}_P$, $\mathbf{R}^{[k,k]} = \mathbf{I}_R$, as in Appendix C. Note that this normalization is arbitrary and thus always possible. We now set out to find all the possible $\mathbf{L}^{[k]}$ not all zero such that (47) holds and \mathbf{P} , \mathbf{R} admissible as in Definition 2.1 on page 2, strictly or relaxed.

The key to the proof is taking the singular value decomposition (SVD) of $\mathbf{L}^{[k]} \in \mathbb{R}^{P \times R}$,

$$\mathbf{L}^{[k]} = \mathbf{U}^{[k]} \mathbf{\Lambda}^{[k]} \mathbf{V}^{[k]\top} \in \mathbb{R}^{P \times R} \quad \forall k, l \quad (48a)$$

$$\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{V}^\top = \bigoplus_{k=1}^K \mathbf{L}^{[k]} \quad , \quad \mathbf{U} = \bigoplus_{k=1}^K \mathbf{U}^{[k]} \quad , \quad \mathbf{\Lambda} = \bigoplus_{k=1}^K \mathbf{\Lambda}^{[k]} \quad , \quad \mathbf{V} = \bigoplus_{k=1}^K \mathbf{V}^{[k]} \quad (48b)$$

where $\mathbf{U}^{[k]} \in \mathbb{R}^{P \times P}$, $\mathbf{V}^{[k]} \in \mathbb{R}^{R \times R}$, and $\mathbf{\Lambda}^{[k]} \in \mathbb{R}^{P \times R}$. Then, (47), which is the key equation in Lemma 2, rewrites as

$$\mathbf{U}^{[k]} \mathbf{\Lambda}^{[k]} \mathbf{V}^{[k]\top} \mathbf{R}^{[k,l]} = \mathbf{P}^{[k,l]} \mathbf{U}^{[l]} \mathbf{\Lambda}^{[l]} \mathbf{V}^{[l]\top} \quad \Rightarrow \quad \mathbf{\Lambda}^{[k]} \underbrace{\mathbf{V}^{[k]\top} \mathbf{R}^{[k,l]} \mathbf{V}^{[l]}}_{\mathbf{B}^{[k,l]}} = \underbrace{\mathbf{U}^{[k]\top} \mathbf{P}^{[k,l]} \mathbf{U}^{[l]}}_{\mathbf{C}^{[k,l]}} \mathbf{\Lambda}^{[l]}$$

$$\mathbf{\Lambda}^{[k]} \mathbf{B}^{[k,l]} = \mathbf{C}^{[k,l]} \mathbf{\Lambda}^{[l]} \in \mathbb{R}^{P \times R} \quad \forall k, l \quad (49a)$$

$$\mathbf{\Lambda} \mathbf{B} = \mathbf{C} \mathbf{\Lambda} \quad (49b)$$

where

$$\mathbb{R}^{P \times P} \ni \mathbf{C}^{[k,l]} \triangleq \mathbf{U}^{[k]\top} \mathbf{P}^{[k,l]} \mathbf{U}^{[l]} \quad \Leftrightarrow \quad \mathbf{P}^{[k,l]} = \mathbf{U}^{[k]} \mathbf{C}^{[k,l]} \mathbf{U}^{[l]\top} \quad , \quad \mathbf{C}^{[k,k]} = \mathbf{I}_P \quad (50a)$$

$$\mathbb{R}^{R \times R} \ni \mathbf{B}^{[k,l]} \triangleq \mathbf{V}^{[k]\top} \mathbf{R}^{[k,l]} \mathbf{V}^{[l]} \quad \Leftrightarrow \quad \mathbf{R}^{[k,l]} = \mathbf{V}^{[k]} \mathbf{B}^{[k,l]} \mathbf{V}^{[l]\top} \quad , \quad \mathbf{B}^{[k,k]} = \mathbf{I}_R \quad (50b)$$

or in matrix form,

$$\mathbf{C} \triangleq \mathbf{U}^\top \mathbf{P} \mathbf{U} \quad \Leftrightarrow \quad \mathbf{P} = \mathbf{U} \mathbf{C} \mathbf{U}^\top \quad (50c)$$

$$\mathbf{B} \triangleq \mathbf{V}^\top \mathbf{R} \mathbf{V} \quad \Leftrightarrow \quad \mathbf{R} = \mathbf{V} \mathbf{B} \mathbf{V}^\top \quad (50d)$$

We note that this transformation does not change the positive-definite nature of \mathbf{C} and \mathbf{B} , as well as all the other admissibility properties that they inherit from \mathbf{P} and \mathbf{R} (Definitions 2.1).

B.1 Non-trivial Solutions to $\mathbf{\Lambda} \mathbf{B} = \mathbf{C} \mathbf{\Lambda}$

The derivation below goes along the same lines as the proof of Schur’s first Lemma, e.g. in [4, Chapter 4]. Note that in Schur’s original lemma, the matrices form a representation; here, this property is replaced by their symmetry. This is essentially the only difference between the proofs. Note that the “relaxed” extension does not exist in the original lemma of Schur nor in the JBD scenario. The key idea of the proof is decoupling (49) into separate equations for \mathbf{B} and \mathbf{C} .

The proof builds on analysing all possible scenarios of nonzero singular values and identifying the admissible cases. Equation (49) is enough to derive conditions associated with the case $P = R$ and $\mathbf{\Lambda}^{[k]} = \lambda \mathbf{I}$, $\lambda \neq 0 \quad \forall k$. However, for all other cases, the equations that we obtained so far are not enough.

Further information can be extracted from (47) by multiplying it on the left or on the right with \mathbf{L}^\top (in fact, this is essentially the same trick as in the proof Schur’s second Lemma in [4, Chapter 4.2] and [35, Lemma A.4], for example ; see also the derivation of the one-dimensional case [10]):

$$\mathbf{L}^\top \mathbf{P} \mathbf{L} = \mathbf{L}^\top \mathbf{L} \mathbf{R} \quad (51a)$$

$$\mathbf{L} \mathbf{R} \mathbf{L}^\top = \mathbf{P} \mathbf{L} \mathbf{L}^\top \quad (51b)$$

Using the fact that the LHS is symmetric,

$$\mathbf{L} \mathbf{L}^\top \mathbf{P} = \mathbf{P} \mathbf{L} \mathbf{L}^\top \quad (52a)$$

$$\mathbf{L}^\top \mathbf{L} \mathbf{R} = \mathbf{R} \mathbf{L}^\top \mathbf{L} \quad (52b)$$

or in blockwise form,

$$\mathbf{L}^{[k]} \mathbf{L}^{[k]\top} \mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]} \mathbf{L}^{[l]} \mathbf{L}^{[l]\top} \quad (52c)$$

$$\mathbf{L}^{[k]\top} \mathbf{L}^{[k]} \mathbf{R}^{[k,l]} = \mathbf{R}^{[k,l]} \mathbf{L}^{[l]} \mathbf{L}^{[l]\top} \quad (52d)$$

where we recall that $\mathbf{P}^{[k,k]}$ and $\mathbf{R}^{[k,k]}$ were normalized such that the $k = l$ blocks do not yield any constraints.

- Equation (52) decouples the two datasets \mathbf{P} and \mathbf{R} . This is the same practice as in the proof of Schur's original second Lemma. The form (52) is the core of Lemma 1, which we now prove.

Using the SVD (48), one can write

$$\mathbf{L}^{[k]} \mathbf{L}^{[k]\top} = \mathbf{U}^{[k]} \mathbf{\Lambda}^{[k]} \underbrace{\mathbf{V}^{[k]\top} \mathbf{V}^{[k]}}_{\mathbf{I}} \mathbf{\Lambda}^{[k]} \mathbf{U}^{[k]\top} = \mathbf{U}^{[k]} \mathbf{\Lambda}_P^{[k]} \mathbf{U}^{[k]\top} \quad (53a)$$

$$\mathbf{L}^{[k]\top} \mathbf{L}^{[k]} = \mathbf{V}^{[k]} \mathbf{\Lambda}^{[k]\top} \underbrace{\mathbf{U}^{[k]\top} \mathbf{U}^{[k]}}_{\mathbf{I}} \mathbf{\Lambda}^{[k]} \mathbf{V}^{[k]\top} = \mathbf{V}^{[k]} \mathbf{\Lambda}_R^{[k]} \mathbf{V}^{[k]\top} \quad (53b)$$

where, without loss of generality,

$$P \geq R > 1 \quad (54)$$

such that

$$\mathbf{\Lambda}^{[k]} = \begin{bmatrix} \lambda_1^{[k]} & & 0 \\ & \ddots & \\ 0 & & \lambda_R^{[k]} \\ \hline & & & \mathbf{0}_{(P-R) \times R} \end{bmatrix} \in \mathbb{R}^{P \times R}. \quad (55)$$

$$\mathbf{\Lambda}_P^{[k]} \triangleq \mathbf{\Lambda}^{[k]} \mathbf{\Lambda}^{[k]\top} = \begin{bmatrix} \lambda_1^{2[k]} & & 0 & | & \mathbf{0}_{R \times (P-R)} \\ & \ddots & & & \\ 0 & & \lambda_R^{2[k]} & | & \\ \hline & & & & \mathbf{0}_{(P-R) \times (P-R)} \end{bmatrix} \in \mathbb{R}^{P \times P} \quad (56a)$$

$$\mathbf{\Lambda}_R^{[k]} \triangleq \mathbf{\Lambda}^{[k]\top} \mathbf{\Lambda}^{[k]} = \begin{bmatrix} \lambda_1^{2[k]} & & 0 \\ & \ddots & \\ 0 & & \lambda_R^{2[k]} \end{bmatrix} \in \mathbb{R}^{R \times R} \quad (56b)$$

Then, (52) can be rewritten as

$$\mathbf{U}^{[k]} \mathbf{\Lambda}_P^{[k]} \mathbf{U}^{[k]\top} \mathbf{P}^{[k,l]} = \mathbf{P}^{[k,l]} \mathbf{U}^{[l]} \mathbf{\Lambda}_P^{[l]} \mathbf{U}^{[l]\top} \Leftrightarrow \mathbf{U} \mathbf{\Lambda}_P \mathbf{U}^\top \mathbf{P} = \mathbf{P} \mathbf{U} \mathbf{\Lambda}_P \mathbf{U}^\top \quad (57a)$$

$$\mathbf{V}^{[k]} \mathbf{\Lambda}_R^{[k]} \mathbf{V}^{[k]\top} \mathbf{R}^{[k,l]} = \mathbf{R}^{[k,l]} \mathbf{V}^{[l]} \mathbf{\Lambda}_R^{[l]} \mathbf{V}^{[l]\top} \Leftrightarrow \mathbf{V} \mathbf{\Lambda}_R \mathbf{V}^\top \mathbf{R} = \mathbf{R} \mathbf{V} \mathbf{\Lambda}_R \mathbf{V}^\top \quad (57b)$$

where $\mathbf{\Lambda}_R = \bigoplus_{k=1}^K \mathbf{\Lambda}_R^{[k]}$ and $\mathbf{\Lambda}_P = \bigoplus_{k=1}^K \mathbf{\Lambda}_P^{[k]}$. Multiplying (57) on the left with \mathbf{U}^\top or \mathbf{V}^\top and on the right with \mathbf{U} or \mathbf{V} , one obtains

$$\mathbf{\Lambda}_P \mathbf{U}^\top \mathbf{P} \mathbf{U} = \mathbf{U}^\top \mathbf{P} \mathbf{U} \mathbf{\Lambda}_P \Leftrightarrow \mathbf{\Lambda}_P^{[k]} \mathbf{U}^{[k]\top} \mathbf{P}^{[k,l]} \mathbf{U}^{[l]} = \mathbf{U}^{[k]\top} \mathbf{P}^{[k,l]} \mathbf{U}^{[l]} \mathbf{\Lambda}_P^{[l]} \quad (58a)$$

$$\mathbf{\Lambda}_R \mathbf{V}^\top \mathbf{R} \mathbf{V} = \mathbf{V}^\top \mathbf{R} \mathbf{V} \mathbf{\Lambda}_R \Leftrightarrow \mathbf{\Lambda}_R^{[k]} \mathbf{V}^{[k]\top} \mathbf{R}^{[k,l]} \mathbf{V}^{[l]} = \mathbf{V}^{[k]\top} \mathbf{R}^{[k,l]} \mathbf{V}^{[l]} \mathbf{\Lambda}_R^{[l]} \quad (58b)$$

Combining (50) with (58) yields

$$\mathbf{\Lambda}_P^{[k]} \mathbf{C}^{[k,l]} = \mathbf{C}^{[k,l]} \mathbf{\Lambda}_P^{[l]} \in \mathbb{R}^{P \times P} \quad (59a)$$

$$\mathbf{\Lambda}_R^{[k]} \mathbf{B}^{[k,l]} = \mathbf{B}^{[k,l]} \mathbf{\Lambda}_R^{[l]} \in \mathbb{R}^{R \times R} \quad (59b)$$

or in matrix form,

$$\mathbf{\Lambda}_P \mathbf{C} = \mathbf{C} \mathbf{\Lambda}_P \in \mathbb{R}^{KP \times KP}, \quad \mathbf{\Lambda}_R \mathbf{B} = \mathbf{B} \mathbf{\Lambda}_R \in \mathbb{R}^{KR \times KR} \quad (59c)$$

Since $\mathbf{C}^{[k,k]} = \mathbf{I}_P$ and $\mathbf{B}^{[k,k]} = \mathbf{I}_R$, the (k, k) th blocks of (49) and (59) do not provide any constraints. Equations (49) and (59) are the simplified and practical counterparts of (52). The advantage of (49) and (59) over (52) is that they are formulated with diagonal, and not block-diagonal, matrices. Therefore, they allow the eigenvalue type of analysis that we present in Appendix B.1.

As we shall soon see, it is useful to rewrite (59) in an entrywise form, as we now explain. The (α, β) th scalar entry of (59a) and the (γ, δ) th scalar entry of (59b) are, for $k \neq l$,

$$\lambda_\alpha^{2[k]} \mathbf{C}_{\alpha\beta}^{[k,l]} = \mathbf{C}_{\alpha\beta}^{[k,l]} \lambda_\beta^{2[l]} \quad , \quad \alpha, \beta \in \{1, \dots, P\} \quad (60a)$$

$$\lambda_\gamma^{2[k]} \mathbf{B}_{\gamma\delta}^{[k,l]} = \mathbf{B}_{\gamma\delta}^{[k,l]} \lambda_\delta^{2[l]} \quad , \quad \gamma, \delta \in \{1, \dots, R\} \quad (60b)$$

Changing sides, (60) can be rewritten as (for $k \neq l$)

$$(\lambda_\alpha^{2[k]} - \lambda_\beta^{2[l]}) \mathbf{C}_{\alpha\beta}^{[k,l]} = 0 \quad , \quad \alpha, \beta \in \{1, \dots, P\} \quad (61a)$$

$$(\lambda_\gamma^{2[k]} - \lambda_\delta^{2[l]}) \mathbf{B}_{\gamma\delta}^{[k,l]} = 0 \quad , \quad \gamma, \delta \in \{1, \dots, R\} \quad (61b)$$

Equation (61) indicates that the relationships between λ within a mixture are not important, only the relationships across mixtures matter; this observation will turn out useful in the analysis of $\mathbf{\Lambda}$, in Appendix B.1. Recall from (56a) that if $P > R$, $\lambda_{\alpha > R}^{[k]} = 0 \forall k$. Then, (61a) can be rewritten as

$$\alpha \leq R, \beta \leq R \quad (\lambda_{\alpha \leq R}^{2[k]} - \lambda_{\beta \leq R}^{2[l]}) \mathbf{C}_{\alpha \leq R, \beta \leq R}^{[k,l]} = 0 \quad (62a)$$

$$\alpha \leq R, \beta > R \quad (\lambda_{\alpha \leq R}^{2[k]} - \underbrace{\lambda_{\beta > R}^{2[l]}}_0) \mathbf{C}_{\alpha \leq R, \beta > R}^{[k,l]} = 0 \quad \text{depends only on } \lambda_{\alpha \leq R}^{2[k]} \quad (62b)$$

$$\alpha > R, \beta \leq R \quad (\underbrace{\lambda_{\alpha > R}^{2[k]}}_0 - \lambda_{\beta \leq R}^{2[l]}) \mathbf{C}_{\alpha > R, \beta \leq R}^{[k,l]} = 0 \quad \text{depends only on } \lambda_{\beta \leq R}^{2[l]} \quad (62c)$$

$$\alpha > R, \beta > R \quad (\underbrace{\lambda_{\alpha > R}^{2[k]} - \lambda_{\beta > R}^{2[l]}}_0) \mathbf{C}_{\alpha > R, \beta > R}^{[k,l]} = 0 \quad \text{no constraints on } \mathbf{C}_{\alpha > R, \beta > R}^{[k,l]} \quad (62d)$$

B.2 Non-trivial Solutions to $\mathbf{\Lambda}_P \mathbf{C} = \mathbf{C} \mathbf{\Lambda}_P$ or $\mathbf{\Lambda}_R \mathbf{B} = \mathbf{B} \mathbf{\Lambda}_R$

The following scenarios summarize the cases where non-zero values of $\mathbf{\Lambda}$ correspond to admissible solutions of $\mathbf{\Lambda}_P \mathbf{C} = \mathbf{C} \mathbf{\Lambda}_P$ or $\mathbf{\Lambda}_R \mathbf{B} = \mathbf{B} \mathbf{\Lambda}_R$, that is, to (62). The derivation is technical; it is based on a detailed analysis of all possible values of the singular values and taking admissibility into account, in full analogy to the derivation of Schur's lemma, see e.g. [4, Chapter 4]. The notation “*” means “no constraint”.

Scenario 3. The first type of non-identifiability is associated with

$$P = R \quad \text{and} \quad \lambda_\alpha^{[k \in \mathcal{D}]} = \lambda \neq 0 \quad \forall \alpha = 1, \dots, R, \quad \mathcal{D} = \{1, \dots, D\}, \quad D \geq 2$$

i.e., entries of λ corresponding to mixtures $\in \mathcal{D}$ are all equal, and different than the rest. This scenario induces

$$\mathbf{C} = \left[\begin{array}{ccc|c} \mathbf{I}_R & \cdots & * & \\ \vdots & \ddots & \vdots & \mathbf{0}_{DR \times (K-D)R} \\ * & \cdots & \mathbf{I}_R & \\ \hline \mathbf{0}_{(K-D)R \times DR} & & & *(K-D)R \times (K-D)R \end{array} \right] = \left[\begin{array}{c|c} \mathbf{C}^{[1:D, 1:D]} & \mathbf{0}_{DR \times (K-D)R} \\ \hline \mathbf{0}_{(K-D)R \times DR} & \mathbf{C}^{[D+1:K, D+1:K]} \end{array} \right]$$

Scenario 4. The second type of non-identifiability is associated with

$$P > R \quad \text{and} \quad \lambda_\alpha^{[k \in \mathcal{D}]} = \lambda \neq 0 \quad \alpha = 1, \dots, R, \quad \mathcal{D} = \{1, \dots, D\}, \quad D \geq 2$$

$$\text{or } P \geq R \quad \text{and} \quad \lambda_\alpha^{[k \in \mathcal{D}]} \neq \lambda_\beta^{[l \in \mathcal{D}]} \neq \lambda_\gamma^{[m \notin \mathcal{D}]} \quad \forall \alpha, \beta, \gamma = 1, \dots, R, \quad D \geq 1$$

i.e., for $P > R$, entries of λ corresponding to at least two mixtures are all equal and different than the rest, or for $P \geq R$, entries of λ corresponding to at least one mixture are all different than each other, and different from the rest. This scenario induces

$$\mathbf{C} = \left[\begin{array}{ccc|c} \mathbf{I}_P & \cdots & \mathbf{0} & \\ \vdots & \ddots & \vdots & \mathbf{0}_{DP \times (K-D)P} \\ \mathbf{0} & \cdots & \mathbf{I}_P & \\ \hline \mathbf{0}_{(K-D)P \times DP} & & & *(K-D)P \times (K-D)P \end{array} \right] = \left[\begin{array}{c|c} \mathbf{I}_D \otimes \mathbf{I}_R & \mathbf{0}_{DP \times (K-D)P} \\ \hline \mathbf{0}_{(K-D)P \times DP} & \mathbf{C}^{[D+1:K, D+1:K]} \end{array} \right]$$

Scenario 5. All other cases of $\mathbf{\Lambda}$ either lead to non-admissible structures (see Section 2.1), or are already subsumed by Scenarios 3–4.

We can reformulate Scenarios 3–4–5 as lemmas.

Lemma 4 (Non-trivial solution to $\mathbf{\Lambda}_P \mathbf{C} = \mathbf{C} \mathbf{\Lambda}_P$, strict). Let $\mathbf{\Lambda}_P^{[k]} = \text{diag}\{\lambda_1^{2[k]}, \dots, \lambda_R^{2[k]}, 0, \dots, 0\}$, $k = 1, \dots, K$ be a set of K diagonal $P \times P$ matrices, with $\lambda_r^{2[k]} \geq 0 \forall k$ and $r = 1, \dots, R$, $P \geq R$. Let $\{\mathbf{C}^{[k,l]}\}_{k,l=1}^K$ be a set of matrices, normalized such that $\mathbf{C}^{[k,k]} = \mathbf{I} \forall k$ and admissible in the strict sense of Definition 2.1. If $\mathbf{\Lambda}_P \mathbf{C} = \mathbf{C} \mathbf{\Lambda}_P$, then $\lambda_r^{2[k]} = \lambda \geq 0 \forall k, r$ with $\lambda > 0$ if $P = R$.

Lemma 5 (Non-trivial solution to $\mathbf{\Lambda}_P \mathbf{C} = \mathbf{C} \mathbf{\Lambda}_P$, relaxed). Let $\mathbf{\Lambda}_P^{[k]} = \text{diag}\{\lambda_1^{2[k]}, \dots, \lambda_R^{2[k]}, 0, \dots, 0\}$, $k = 1, \dots, K$ be a set of K diagonal $P \times P$ matrices, with $\lambda_r^{2[k]} \geq 0 \forall k$ and $r = 1, \dots, R$, $P \geq R$. Let $\{\mathbf{C}^{[k,l]}\}_{k,l=1}^K$ be a set of matrices, normalized such that $\mathbf{C}^{[k,k]} = \mathbf{I} \forall k$ and admissible in the relaxed sense of Definition 2.1. If $\mathbf{\Lambda}_P \mathbf{C} = \mathbf{C} \mathbf{\Lambda}_P$, then we can make the following statements about the values of $\mathbf{\Lambda}$:

1. $\lambda_r^{2[k]} = \lambda \geq 0 \forall r, k \in \mathcal{D}$, $\mathcal{D} = \{1, \dots, D\}$, $D \geq 2$, if

$$(a) \mathbf{C} \text{ has structure } \left[\begin{array}{c|c} \mathbf{C}^{[1:D,1:D]} & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & \mathbf{C}^{[D+1:K,D+1:K]} \end{array} \right] \text{ and } P = R, \text{ or}$$

$$(b) \mathbf{C} \text{ has structure } \left[\begin{array}{c|c} \mathbf{I}_D \otimes \mathbf{I}_R & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & \mathbf{C}^{[D+1:K,D+1:K]} \end{array} \right] \text{ and } P > R.$$

2. $\lambda_\alpha^{[k \in \mathcal{D}]} \neq \lambda_\beta^{[l \in \mathcal{D}]} \neq \lambda_\gamma^{[m \notin \mathcal{D}]} \forall \alpha, \beta, \gamma = 1, \dots, R$, $\mathcal{D} = \{1, \dots, D\}$, $D \geq 1$, if \mathbf{C} has structure

$$\left[\begin{array}{c|c} \mathbf{I}_D \otimes \mathbf{I}_R & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & \mathbf{C}^{[D+1:K,D+1:K]} \end{array} \right] \text{ and } P \geq R.$$

3. All other non-zero options of $\mathbf{\Lambda}$ either correspond to non-admissible \mathbf{C} , or are subsumed by the above two options by reordering the indices k .

B.3 Completing the Proof

In order to complete the proof, we have to form a link between Scenarios 3–4–5 on page 15 or Lemma 4–5, which were derived for $\mathbf{\Lambda}_P \mathbf{C} = \mathbf{C} \mathbf{\Lambda}_P$ and $\mathbf{\Lambda}_R \mathbf{B} = \mathbf{C} \mathbf{\Lambda}_R$ and in which \mathbf{B} and \mathbf{C} were decoupled, via $\mathbf{\Lambda} \mathbf{B} = \mathbf{B} \mathbf{\Lambda}$ (49). The result is summarized in the following Scenarios 6–7–8.

Scenario 6. From Scenario 3, where

$$P = R \quad \text{and} \quad \lambda_\alpha^{[k \in \mathcal{D}]} = \lambda \neq 0 \quad \forall \alpha = 1, \dots, R, \quad \mathcal{D} = \{1, \dots, D\}, \quad D \geq 2$$

we obtain that both \mathbf{B} and \mathbf{C} have the structure

$$\mathbf{B} = \left[\begin{array}{c|c} \mathbf{B}^{[1:D,1:D]} & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & \mathbf{B}^{[D+1:K,D+1:K]} \end{array} \right], \quad \mathbf{C} = \left[\begin{array}{c|c} \mathbf{C}^{[1:D,1:D]} & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & \mathbf{C}^{[D+1:K,D+1:K]} \end{array} \right]$$

adding the constraint $\mathbf{\Lambda}_R \mathbf{B} = \mathbf{C} \mathbf{\Lambda}_R$ implies that, for $\{k, l\} \in \mathcal{D}$, $\lambda \mathbf{I} \mathbf{B}^{[k,l]} = \mathbf{C}^{[k,l]} \lambda \mathbf{I}$. Since $\lambda \neq 0$, we obtain

$$\mathbf{B}^{[k \in \mathcal{D}, l \in \mathcal{D}]} = \mathbf{C}^{[k \in \mathcal{D}, l \in \mathcal{D}]}$$

Scenario 7. From Scenario 4, where

$$P > R \quad \text{and} \quad \lambda_\alpha^{[k \in \mathcal{D}]} = \lambda \neq 0 \quad \alpha = 1, \dots, R, \quad \mathcal{D} = \{1, \dots, D\}, \quad D \geq 2$$

$$\text{or } P \geq R \quad \text{and} \quad \lambda_\alpha^{[k \in \mathcal{D}]} \neq \lambda_\beta^{[l \in \mathcal{D}]} \neq \lambda_\gamma^{[m \notin \mathcal{D}]} \quad \forall \alpha, \beta, \gamma = 1, \dots, R, \quad D \geq 1$$

the equality $\mathbf{\Lambda}_R \mathbf{B} = \mathbf{C} \mathbf{\Lambda}_R$, together with relaxed admissibility constraints in Definition 2.1, eventually leads to the conclusion that the only corresponding admissible structures are

$$\mathbf{B} = \left[\begin{array}{c|c} \mathbf{I}_D \otimes \mathbf{I}_R & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & \mathbf{B}^{[D+1:K,D+1:K]} \end{array} \right], \quad \mathbf{C} = \left[\begin{array}{c|c} \mathbf{I}_D \otimes \mathbf{I}_P & \mathbf{0}_{DP \times (K-D)P} \\ \mathbf{0}_{(K-D)P \times DP} & \mathbf{C}^{[D+1:K,D+1:K]} \end{array} \right]$$

Scenario 8. All other cases of $\mathbf{\Lambda}$ either lead to non-admissible structures (see Section 2.1), or are already subsumed by Scenarios 1–2.

Scenarios 6–7–8 lead to the following two lemmas.

Lemma 6 (Non-trivial solution to $\mathbf{\Lambda B} = \mathbf{C A}$, strict). Let $\{\mathbf{B}^{[k,l]}\}_{k,l=1}^K$ and $\{\mathbf{C}^{[k,l]}\}_{k,l=1}^K$ be two sets of matrices, of dimensions $R \times R$ and $P \times P$, respectively, normalized such that $\mathbf{B}^{[k,k]} = \mathbf{I}_R$ and $\mathbf{C}^{[k,k]} = \mathbf{I}_P \forall k$ and admissible in the strict sense of Definition 2.1. Without loss of generality, $P \geq R$. Let $\mathbf{\Lambda}^{[k]} = \begin{bmatrix} \text{diag}\{\lambda_1^{[k]}, \dots, \lambda_R^{[k]}\} \\ \mathbf{0}_{(P-R) \times R} \end{bmatrix}$, $k = 1, \dots, K$ be a set of K $P \times R$ matrices, such that $\mathbf{\Lambda B} = \mathbf{C A}$. Then, either $\mathbf{\Lambda} = \mathbf{0}$, or $P = R$, $\mathbf{\Lambda} = \lambda \mathbf{I}$, $\lambda \neq 0$ and $\mathbf{B} = \mathbf{C}$.

Lemma 7 (Non-trivial solution to $\mathbf{\Lambda B} = \mathbf{C A}$, relaxed). Let $\{\mathbf{B}^{[k,l]}\}_{k,l=1}^K$ and $\{\mathbf{C}^{[k,l]}\}_{k,l=1}^K$ be two sets of matrices, of dimensions $R \times R$ and $P \times P$, respectively, normalized such that $\mathbf{B}^{[k,k]} = \mathbf{I}_R$ and $\mathbf{C}^{[k,k]} = \mathbf{I}_P \forall k$ and admissible in the relaxed sense of Definition 2.1. Without loss of generality, $P \geq R$. Let $\mathbf{\Lambda}^{[k]} = \begin{bmatrix} \text{diag}\{\lambda_1^{[k]}, \dots, \lambda_R^{[k]}\} \\ \mathbf{0}_{(P-R) \times R} \end{bmatrix}$, $k = 1, \dots, K$ be a set of K $P \times R$ matrices. If $\mathbf{\Lambda B} = \mathbf{C A}$, then we can make the following statements about the values of $\mathbf{\Lambda}$, \mathbf{B} and \mathbf{C} :

1. $\lambda_r^{2[k]} = \lambda \geq 0 \forall r, k \in \mathcal{D}$, $\mathcal{D} = \{1, \dots, D\}$, $D \geq 2$, if

(a) $P = R$, \mathbf{B} and \mathbf{C} have the same structure, $\begin{bmatrix} \mathbf{B}^{[1:D,1:D]} & | & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & | & \mathbf{B}^{[D+1:K,D+1:K]} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{C}^{[1:D,1:D]} & | & \mathbf{0}_{DP \times (K-D)P} \\ \mathbf{0}_{(K-D)P \times DP} & | & \mathbf{C}^{[D+1:K,D+1:K]} \end{bmatrix}$, respectively, $\mathbf{B}^{[1:D,1:D]} = \mathbf{C}^{[1:D,1:D]}$, i.e. the upper left block is identical, or

(b) $P > R$, \mathbf{B} and \mathbf{C} have the same structure $\begin{bmatrix} \mathbf{I}_D \otimes \mathbf{I}_R & | & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & | & \mathbf{C}^{[D+1:K,D+1:K]} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{I}_D \otimes \mathbf{I}_P & | & \mathbf{0}_{DP \times (K-D)P} \\ \mathbf{0}_{(K-D)P \times DP} & | & \mathbf{B}^{[D+1:K,D+1:K]} \end{bmatrix}$, respectively.

2. $\lambda_\alpha^{[k \in \mathcal{D}]} \neq \lambda_\beta^{[l \in \mathcal{D}]} \neq \lambda_\gamma^{[m \notin \mathcal{D}]}$, $k \neq l \neq m$, $\forall \alpha, \beta, \gamma = 1, \dots, R$, $\mathcal{D} = \{1, \dots, D\}$, $D \geq 1$, if $P \geq R$ and \mathbf{B} and \mathbf{C} have structure $\begin{bmatrix} \mathbf{I}_D \otimes \mathbf{I}_R & | & \mathbf{0}_{DR \times (K-D)R} \\ \mathbf{0}_{(K-D)R \times DR} & | & \mathbf{C}^{[D+1:K,D+1:K]} \end{bmatrix}$ and $\begin{bmatrix} \mathbf{I}_D \otimes \mathbf{I}_P & | & \mathbf{0}_{DP \times (K-D)P} \\ \mathbf{0}_{(K-D)P \times DP} & | & \mathbf{B}^{[D+1:K,D+1:K]} \end{bmatrix}$, respectively.

3. All other non-zero options of $\mathbf{\Lambda}$ either correspond to non-admissible \mathbf{B} or \mathbf{C} , or are subsumed by the above two options by reordering the indices k .

Applying Lemma 6–7 to (49) concludes our proof.

C Normalization

The normalization scheme that we use is as follows.

$$\mathbf{P} = \mathbf{\Omega}_{ii} \mathbf{S}_{ii} \mathbf{\Omega}_{ii}^\top \Leftrightarrow \mathbf{S}_{ii} = \mathbf{\Omega}_{ii}^{-1} \mathbf{P} \mathbf{\Omega}_{ii}^{-\dagger} \quad (63a)$$

$$\mathbf{R} = \mathbf{\Omega}_{jj} \mathbf{S}_{jj} \mathbf{\Omega}_{jj}^\top \Leftrightarrow \mathbf{S}_{jj} = \mathbf{\Omega}_{jj}^{-1} \mathbf{R} \mathbf{\Omega}_{jj}^{-\dagger} \quad (63b)$$

or in blockwise form,

$$\mathbf{P}^{[k,l]} = \mathbf{\Omega}_{ii}^{[k]} \mathbf{S}_{ii}^{[k,l]} \mathbf{\Omega}_{ii}^{[l]\top} \in \mathbb{R}^{m_i \times m_i} \quad (63c)$$

$$\mathbf{R}^{[k,l]} = \mathbf{\Omega}_{jj}^{[k]} \mathbf{S}_{jj}^{[k,l]} \mathbf{\Omega}_{jj}^{[l]\top} \in \mathbb{R}^{m_j \times m_j} \quad (63d)$$

such that

$$\mathbf{P}^{[k,k]} = \mathbf{I}_{m_i} \quad (64a)$$

$$\mathbf{R}^{[k,k]} = \mathbf{I}_{m_j}, \quad (64b)$$

where

$$\mathbf{\Omega}_{ii}^{[k]} \triangleq (\mathbf{S}_{ii}^{[k,k]})^{-\frac{1}{2}} \in \mathbb{S}_+^{m_i \times m_i}, \quad \mathbf{\Omega}_{ii} = \bigoplus_{k=1}^K \mathbf{\Omega}_{ii}^{[k]} \quad (65)$$

References

- [1] M. W. Kirson, *Introductory Algebra for Physicists*, 2013, ch. Irreducible representations.
- [2] C. W. Curtis, *Pioneers of Representation Theory: Frobenius, Burnside, Schur, and Brauer*, ser. History of Mathematics. Providence, RI, USA: American Mathematical Society, 1999, vol. 15.
- [3] A. Alexanderian, “A basic note on group representations and Schurs lemma,” <http://www4.ncsu.edu/~aalexan3/articles/schur.pdf>.
- [4] D. Vvedensky, “Group theory course notes,” <http://www.cmth.ph.ic.ac.uk/people/d.vvedensky/courses.html>, 2001.
- [5] I. Schur, *Neue Begründung der Theorie der Gruppencharaktere*, ser. Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften zu Berlin, 1905. [Online]. Available: <https://books.google.fr/books?id=QoUWSQAACAAJ>
- [6] D. Lahat, J.-F. Cardoso, and H. Messer, “Identifiability of second-order multidimensional ICA,” in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 1875–1879.
- [7] L. De Lathauwer, “Decompositions of a higher-order tensor in block terms. Part II: Definitions and uniqueness,” *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1033–1066, 2008.
- [8] D. Lahat and C. Jutten, “Joint blind source separation of multidimensional components: Model and algorithm,” in *Proc. EUSIPCO*, Lisbon, Portugal, Sep. 2014, pp. 1417–1421.
- [9] —, “Joint independent subspace analysis using second-order statistics,” GIPSA-Lab, Grenoble, France, Technical report hal-01132297, Mar. 2015, accepted for publication with mandatory minor revisions (AQ) in the IEEE Transactions on Signal Processing.
- [10] —, “An alternative proof for the identifiability of independent vector analysis using second order statistics,” in *Proc. ICASSP*, Shanghai, China, Mar. 2016.
- [11] M. Congedo, R. Phlypo, and J. Chatel-Goldman, “Orthogonal and non-orthogonal joint blind source separation in the least-squares sense,” in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 1885–1889.
- [12] J. Chatel-Goldman, M. Congedo, and R. Phlypo, “Joint BSS as a natural analysis framework for EEG-hyperscanning,” in *Proc. ICASSP*, Vancouver, Canada, May 2013, pp. 1212–1216.
- [13] R. F. Silva, S. Plis, T. Adalı, and V. D. Calhoun, “Multidataset independent subspace analysis extends independent vector analysis,” in *Proc. ICIP*, Paris, France, Oct. 2014, pp. 2864–2868.
- [14] T. Adalı, M. Anderson, and G.-S. Fu, “Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging,” *IEEE Signal Process. Mag.*, pp. 18–33, May 2014.
- [15] M. Sørensen and L. De Lathauwer, “Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(L_{r,n}, L_{r,n}, 1)$ terms—part I: Uniqueness,” *SIAM J. Matrix Anal. Appl.*, vol. 36, no. 2, pp. 496–522, Apr. 2015.
- [16] D. Lahat, T. Adalı, and C. Jutten, “Multimodal data fusion: An overview of methods, challenges and prospects,” *Proc. IEEE*, vol. 103, no. 9, pp. 1449–1477, Sep. 2015.
- [17] T. Maehara and K. Murota, “Algorithm for error-controlled simultaneous block-diagonalization of matrices,” *SIAM J. Matrix Anal. Appl.*, vol. 32, no. 2, pp. 605–620, 2011.
- [18] —, “Error-controlling algorithm for simultaneous block-diagonalization and its application to independent component analysis,” *JSIAM Letters*, vol. 2, pp. 131–134, 2010.
- [19] —, “A numerical algorithm for block-diagonal decomposition of matrix *-algebras with general irreducible components,” *Japan Journal of Industrial and Applied Mathematics*, vol. 27, no. 2, pp. 263–293, Sep. 2010.
- [20] H. W. Gutch, T. Maehara, and F. J. Theis, “Second order subspace analysis and simple decompositions,” in *Latent Variable Analysis and Signal Separation*, ser. LNCS, V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, Eds., vol. 6365. Heidelberg: Springer, 2010, pp. 370–377.
- [21] D. Lahat and C. Jutten, “On the uniqueness of coupled matrix block diagonalization in the joint analysis of multiple datasets (talk),” in *SIAM Conference on Applied Linear Algebra*, Atlanta, GA, USA, Oct. 2015, <http://meetings.siam.org/sess/dsp.talk.cfm?p=72077>.
- [22] T. W. Anderson, *An introduction to multivariate statistical analysis*. John Wiley & Sons, 1958.
- [23] D. Lahat, J.-F. Cardoso, and H. Messer, “Joint block diagonalization algorithms for optimal separation of multidimensional components,” in *Latent Variable Analysis and Signal Separation*, ser. LNCS, F. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, Eds., vol. 7191. Springer, 2012, pp. 155–162.
- [24] M. Anderson, “Independent vector analysis: Theory, algorithms, and applications,” Ph.D. dissertation, Graduate School of the University of Maryland, Baltimore County, 2013.
- [25] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adalı, “Independent vector analysis: Identification conditions and performance bounds,” *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4399–4410, Sep. 2014.
- [26] J. R. Magnus and H. Neudecker, “The commutation matrix: Some properties and applications,” *Ann. Statist.*, vol. 7, no. 2, pp. 381–394, Mar. 1979.

- [27] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, ser. Mathematics and its Applications. Chichester, West Sussex, England: Ellis Horwood Limited, 1981.
- [28] K. B. Petersen and M. S. Pedersen, “The matrix cookbook,” Oct. 2008, version 20081110. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?3274>
- [29] A. Cichocki, D. Mandic, A. H. Phan, C. Caiafa, G. Zhou, Q. Zhao, and L. De Lathauwer, “Tensor decompositions for signal processing applications: From two-way to multiway component analysis,” *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 145–163, Mar. 2015.
- [30] J. W. Brewer, “Kronecker products and matrix calculus in system theory,” *IEEE Trans. Circuits Syst.*, vol. CAS-25, no. 9, pp. 772–781, Sep. 1978.
- [31] S. Liu and G. Trenkler, “Hadamard, Khatri-Rao, Kronecker and other matrix products,” *Int. J. Inform. Syst. Sci.*, vol. 4, no. 1, pp. 160–177, 2008.
- [32] R. Bro, “Multi-way analysis in the food industry: models, algorithms, and applications,” Ph.D. dissertation, Chemometrics Group, Food Technology. Department of Dairy and Food Science. Royal Veterinary and Agricultural University, Denmark, 1998.
- [33] C. R. Rao and S. K. Mitra, *Generalized inverse of matrices and its applications*. John Wiley & Sons, 1971, vol. 7.
- [34] R. P. McDonald, “A simple comprehensive model for the analysis of covariance structures: Some remarks on applications,” *Br. J. Math. Stat. Psychol.*, vol. 33, no. 2, pp. 161–183, 1980.
- [35] K. Murota, Y. Kanno, M. Kojima, and S. Kojima, “A numerical algorithm for block-diagonal decomposition of matrix *-algebras with application to semidefinite programming,” *Japan J. Indust. Appl. Math.*, vol. 27, no. 1, pp. 125–160, Jun. 2010.