Joint Independent Subspace Analysis: Uniqueness and Identifiability

Dana Lahat, Christian Jutten

Joint Independent Subspace Analysis: Uniqueness and Identifiability

Dana Lahat and Christian Jutten, *Fellow, IEEE*

***Abstract*—This paper deals with the identifiability of joint independent subspace analysis (JISA). JISA is a recently-proposed framework that subsumes independent vector analysis (IVA) and independent subspace analysis (ISA). Each underlying mixture can be regarded as a dataset; therefore, JISA can be used for data fusion. In this paper, we assume that each dataset is an overdetermined mixture of several multivariate Gaussian processes, each of which has independent and identically distributed samples. This setup is not identifiable when each mixture is considered individually. Given these assumptions, JISA can be restated as coupled block diagonalization (CBD) of its correlation matrices. Hence, JISA identifiability is tantamount to CBD uniqueness. In this work, we provide necessary and sufficient conditions for uniqueness and identifiability of JISA and CBD. Our analysis is based on characterizing all the cases in which the Fisher information matrix is singular. We prove that non-identifiability may occur only due to pairs of underlying random processes with the same dimension. Our results provide further evidence that irreducibility has a central role in the uniqueness analysis of block-based decompositions. Our contribution extends previous results on the uniqueness and identifiability of ISA, IVA, coupled matrix and tensor decompositions. We provide examples to illustrate our results.**

***Index Terms*—Blind source separation, independent vector analysis, block decompositions, uniqueness, identifiability, coupled decompositions, data fusion.**

I. Introduction

THIS theoretical paper deals with the identifiability of joint independent subspace analysis (JISA) [1]–[3]. JISA is a recently-proposed model that extends independent subspace analysis (ISA) [4], [5] by considering several different ISA problems that are linked by statistical dependencies among the latent multivariate random processes. Another way of looking at JISA is as an extension of independent vector analysis (IVA) [6] to mixtures of multivariate, instead of univariate, random variables. Both ISA and IVA are themselves extensions of independent component analysis (ICA) [7], a simple yet powerful concept that has given rise to the very vast domain of blind source separation (BSS) [8], [9]. As such, JISA is a rich framework that subsumes the versatility of the models that it is inspired from.

JISA is a very general framework that is able to exploit any of the types of diversity that are traditionally used in single-mixture BSS, such as complex-valued data, higher-order statistics (HOS), sample non-stationarity, and dependence among samples, to name a few [1], [3], [10]. Due to the presence of multiple mixtures (which may be interpreted as multiple datasets), and the links among them, the algebraic structures associated with JISA are more elaborate than the ones associated with single-mixture BSS; see, e.g., [1, Section VI], [10], for concrete examples. Algorithms for JISA under various model assumptions are described, e.g., in [1]–[3], [10], [11]. Often, each type of diversity embedded in the model further enhances identifiability (e.g., [12]).

In this paper, however, we choose to focus on a JISA model that does not take into account any of these traditional types of diversity. In this model, each of the underlying multivariate random processes, in each mixture, is Gaussian with independent and identically distributed (i.i.d.) samples. From a theoretical point of view, this setup is of particular interest because each underlying ISA problem is not identifiable individually. In this case, its identifiability, if exists, is due to the link among mixtures. Therefore, this JISA setup allows us to isolate the added value of the link between datasets [12], [13]. Understanding the transition point between non-identifiability and identifiability can provide us with better understanding of coupled decompositions and data fusion. The results that we present in this paper add to the increasing amount of theoretical evidence (e.g., [10], [14]–[16]) that the special type of diversity associated with data fusion is sufficient to obtain identifiability of models that are otherwise not identifiable. From a practical point of view, assuming Gaussian i.i.d. variables means that second-order statistics (SOS) are sufficient, and that no temporal properties between samples, or non-Gaussianity, are required, or used.

Previous work on this JISA model has dealt with algorithms (e.g., [2], [11]) and small error analysis, including the derivation of Cramér-Rao lower bound (CRLB) and Fisher information matrix (FIM) [1]. This paper deals with its uniqueness and identifiability. In previous work, we obtained supporting evidence that this JISA model can be identifiable. First, we have shown in [1] that the data (observations) always provide a sufficient number of constraints with respect to (w.r.t.) the number of unknowns in the model, as soon as the following additional simplifying assumptions hold: (i) all mixing matrices have full column rank, and (ii) the dimensions of the corresponding latent multivariate random processes are identical in all datasets; within each mixture, the different multivariate random processes may have different dimensions. We shall formulate these assumptions mathematically in Section II. Second, in [1], [11], given these model assumptions, we derived a closed-form expression for the FIM, and demonstrated its invertibility numerically, for randomly generated data.

D. Lahat is with IRIT, Université de Toulouse, CNRS, Toulouse, France (email: Dana.Lahat@irit.fr). This work was carried out when D. Lahat was with GIPSA-lab, Grenoble, France.

C. Jutten is with Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France (email: Christian.Jutten@gipsa-lab.grenoble-inp.fr).

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

In this paper, however, we derive the identifiability of this JISA model rigorously, by characterizing all the cases in which this model is not identifiable, and showing that non-identifiability occurs only in very particular cases. Previous theoretical results on the identifiability of SOS-based IVA in [14], [17], [18], and SOS-based nonstationary ISA [19] thus become special cases of the new results in this paper. Assumption (i) implies that the identifiability of this model depends only on the SOS of the latent random processes and not on the mixing matrices. In [1], this assumption allowed us to write the CRLB and FIM using closed form and analytically tractable expressions. It is worth mentioning that this JISA model may be identifiable even if assumption (i) is not satisfied, in certain cases, as indicated by related results in [16], [20]. As for assumption (ii), it is important to note that although, in many cases, the same conclusions apply also when assumption (ii) is relaxed, this is not always the case, as demonstrated in [21].

JISA is a statistically motivated model. However, it can also be considered from an algebraic point of view, as we now explain. As shown in [1], [11], the JISA model that we have just described can be reformulated as an (approximate) coupled block diagonalization (CBD) of the (sample) covariance matrices of the observations, if they exist and have finite values. These covariance matrices are sufficient statistics for full model identifiability. Therefore, in this case, JISA identifiability amounts to CBD uniqueness. CBD is formulated mathematically in Section II. The fact that our model is tantamount to a reduction of an ensemble of matrices to a block-diagonal form implies that we have to characterize mathematically the fact that these diagonal blocks cannot be further factorized into smaller ones. In this paper, we show that this property, called irreducibility, plays a central part in the identifiability of our model.

Our identifiability analysis is based on characterizing all the cases in which the FIM is singular [22]. As mentioned earlier, this FIM was derived in closed-form in [1]. As we show in Appendix B, our analysis boils down to characterizing the set of non-trivial solutions to a system of coupled matrix equations, given rank and irreducibility constraints on its coefficients. This problem does not exist in the literature, and thus addressing it is another contribution arising from this work. In Appendix A, we provide new identities on partitioned matrices. These identities resulted from our analysis of the FIM, whose structure involves Khatri-Rao products of block partitioned matrices. The derivation of the solutions to this constrained system of matrix equations is explained in detail in [23], [24]. In this paper, we only cite the relevant results.

A similar FIM-based approach was used in [19], in the identifiability analysis of ISA of piecewise-stationary overdetermined Gaussian multivariate processes. Similarly to JISA and CBD, this ISA model can be represented algebraically, through joint block diagonalization (JBD) of its correlation matrices [25]. Therefore, ISA identifiability can be recast in algebraic terms of JBD uniqueness. JBD differs from CBD in the number of transformations that block-diagonalize the observations: one in JBD, several in CBD. In Section VIII-A, we show that these models are not only similar but in fact, in certain respects, ISA and JBD can be regarded as special cases of JISA and CBD. The uniqueness and identifiability results for these ISA and JBD models, in [19], [26], [27], are based on Schur's lemma on irreducible representations [28], whose link with JBD was first pointed out in [29]. Schur's lemma deals with a single transformation applied to an ensemble of matrices, and is thus not applicable to CBD. In this paper, we extend the concept of irreducibility to multiple transformations, and show that it is equally crucial in analysing JISA and CBD uniqueness and identifiability. In Section VIII, we discuss some of the links between the results in [19], [26], [27] and those in this paper, and explain how the results in this paper can be regarded as a generalization of the former. CBD and JBD will be described more rigorously in the upcoming sections of this paper.

Our discussion of JISA is motivated by its ability to provide a flexible framework for coupled processing of multiple datasets, e.g.: (i) recordings of one scene by different devices, (ii) recordings of similar scenes (or subjects) with the same device, (iii) recordings of one scene (or subject) with one device at different time windows. In a data fusion context, each underlying ISA mixture represents a dataset. The statistical links among mixtures amount to links among datasets. In the literature, and in particular in the context of data fusion, this type of links among datasets is sometimes referred to as "soft". Soft links allow each dataset in the ensemble to remain in its most explanatory form, with its own parameters, and thus allow a high degree of flexibility in fusing heterogeneous datasets. This is the case in JISA, where each dataset has its own mixing matrix and statistical model for the signals. The other option is "hard" links, where datasets deterministically share some of their factors. We refer the reader to [13] for further discussion of this matter.

The *raison d'être* of the ISA aspect of JISA is that in various real-world applications, the assumption of classical BSS, that each signal can be modeled by a single random variable, is too restrictive. Although univariate methods are sometimes used to achieve ISA, it has been shown that true multivariate methods enhance accuracy and interpretability of the output (e.g., [30]). In addition, algorithms based on this idea avoid futile attempts to further factorize irreducible subspaces, and thus improve computational efficiency [4, Section 8] [11]. These ideas extend naturally to JISA. JISA is the first ICA-type framework to be able to exploit multidimensional block structures in an ensemble of linked datasets.

The data fusion capacities of JISA are inherited from IVA. The original motivation for IVA was dealing with the arbitrary permutation of the estimated spectral coefficients in frequency domain BSS of convolutive mixtures of acoustic sources [6]. This property extends naturally to JISA. Recently, IVA has shown useful for a broad range of applications. For example, as a framework for flexible modeling of the signal subspace that is common to multiple subjects, as opposed to earlier state-of-the-art ICA-based methods that are more rigid [31]. Kim et al. [31] emphasize the importance of preserving subject variability in multi-subject functional magnetic resonance imaging (fMRI) analyses, for various personalized prediction tasks, for example.

An ISA-based approach has already been found useful in various applications, including electrocardiography (ECG) [5], [32], fMRI [33], electroencephalography (EEG) [34], astrophysics [35], [36], and separation of mixed audio sources [37]. These references also propose different strategies for determining the most useful size of the subspaces. Based on these properties, Silva et al. [3] suggest that the flexibility of JISA in considering multiple datasets with heterogeneous structures has great potential for multimodal neuroimaging–genetics data analysis, e.g., for combining information in fMRI spatial maps with genetic single nucleotide polymorphism (SNP) arrays.

Biomedical, astrophysical, and audio data, which we have just mentioned within the context of ISA, are often studied within a data fusion framework [13]. It is thus natural to suggest JISA as a framework that can capture both their multiset and within-dataset structures. The JISA model that we consider in this paper is applicable to any data that admit a JISA model (possibly with more types of diversity), and whose SOS satisfy our model assumptions, as well as the identifiability conditions that we derive in this paper. Naturally, for any real-life data, these requirements will, at best, be satisfied only approximately. For example, a potential application of JISA is the analysis of ECG data, whose multidimensional nature was demonstrated in [32], for fetal monitoring. Li et al. [38] showed that this task can be achieved by IVA, in a suboptimal "mismodeling" approach [30]. Given the ongoing interest in non-invasive methods for fetal monitoring, in multi-subject and multimodal setups (see, e.g., [39] and references therein), JISA is a potentially useful candidate to address some of these challenges. Our preliminary experimental results on the ECG data of [32], using an approach similar to that in [38], indicate that the fetal electrocardiogram (FECG) can be extracted to a reasonable extent even by the simplest JISA model that we consider in this paper, using the algorithm in [11]. Due to lack of space, we omit the details. Despite the theoretically attractive advantages of data fusion and coupled decompositions, when handling real-world data, there are many more considerations to take into account, as discussed, e.g., in [13], [40], [41]. Therefore, it is of utmost importance to first understand the theoretical limitations of the basic model before encumbering it with additional complicating factors.

This work is the first to deal theoretically with the identifiability of JISA and the uniqueness of the CBD associated with it. Until now, uniqueness results for coupled decompositions addressed only models in which the link between datasets was through a one-dimensional or a rank-1 elements; e.g., coupled block term decomposition (BTD) [15] and IVA [14], [16]–[18]). In this paper, however, the link among datasets is via multivariate statistics and terms of rank larger than one.

The results in this manuscript have previously been presented orally in [42]–[44] and in a technical report [24], however, they have never been published.

A. Notations

The following notations will be used throughout this paper. Symbols $\cdot^{\top}$, $\cdot^{-\top}$, $\cdot^{\mathsf{H}}$ denote transpose, inverse transpose, and conjugate transpose (Hermitian), respectively. $\mathrm{E}\{\cdot\}$ denotes expectation. $\mathbf{I}_M$ and $\mathbf{0}_{M\times N}$ denote an identity and an all-zero matrices, respectively. $\mathbf{1}_K$ denotes an all-ones vector of length K. Scalars, vectors, matrices, and sets, are denoted by a, $\mathbf{a}$, $\mathbf{A}$, and $\mathcal{A}$, respectively. All vectors are column vectors, unless stated otherwise. $\mathbf{A}_i$ is the ith column block of a column-wise partitioned block matrix $\mathbf{A}$. $\mathbf{A}_{ij}$ is the (i,j)th block of a partitioned block matrix $\mathbf{A}$. Vector $\mathbf{a}_i$ denotes the ith sub-vector of a partitioned vector $\mathbf{a}$, or the ith column of matrix $\mathbf{A}$. a_i and a_{ij} denote the ith and (i,j)th scalar elements of $\mathbf{a}$ and $\mathbf{A}$, respectively. $\mathbf{A}^{[k]}$ and $\mathbf{a}^{[k]}$ denote a matrix $\mathbf{A}$ or a vector $\mathbf{a}$ indexed by k. Notations $\mathbf{A}^{[k]\top}$, $\mathbf{A}^{-[k]}$ and $\mathbf{A}^{-[k]\top}$ should be read as $(\mathbf{A}^{[k]})^{\top}$, $(\mathbf{A}^{[k]})^{-1}$ and $(\mathbf{A}^{[k]})^{-\top}$, respectively. $\mathbf{M}^{[k\in\mathcal{D}]}$ stands for "$\mathbf{M}^{[k]}$ for all $k\in\mathcal{D}$". For double-indexed matrices, $\mathbf{M}^{[1:K,1:L]} \triangleq \begin{bmatrix} \mathbf{M}^{[1,1]} & \cdots & \mathbf{M}^{[1,L]} \\ \vdots & & \vdots \\ \mathbf{M}^{[K,1]} & \cdots & \mathbf{M}^{[K,L]} \end{bmatrix}$. A direct sum of K matrices is denoted by $\oplus_{k=1}^{K}\mathbf{N}^{[k]} = \mathbf{N}^{[1]}\oplus\cdots\oplus\mathbf{N}^{[K]} = \begin{bmatrix} \mathbf{N}^{[1]} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{N}^{[K]} \end{bmatrix}$. $\mathrm{span}(\mathbf{A})$ denotes the column space of $\mathbf{A}$. $\mathrm{vec}(\mathbf{X})$ stacks the entries of $\mathbf{X}$ in one vector, column-wise. We use the following operations on partitioned matrices:

Product name	Notation
Kronecker	$\otimes$
Hadamard	$*$
Khatri-Rao column-wise	$\odot$
Khatri-Rao block-column-wise	$\boxed{\vert}$
Khatri-Rao for partitioned matrices	$\boxplus$

where $\mathbf{A}\otimes\mathbf{B} \triangleq \begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$, $\mathbf{A}*\mathbf{B} \triangleq \begin{bmatrix} a_{11}b_{11} & a_{12}b_{12} & \cdots \\ a_{21}b_{21} & a_{22}b_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$, $\mathbf{A}\odot\mathbf{B} \triangleq \begin{bmatrix}\mathbf{a}_1\otimes\mathbf{b}_1 & \mathbf{a}_2\otimes\mathbf{b}_2 & \cdots\end{bmatrix}$, $\mathbf{A} \boxed{\vert} \mathbf{B} \triangleq \begin{bmatrix}\mathbf{A}_1\otimes\mathbf{B}_1 & \mathbf{A}_2\otimes\mathbf{B}_2 & \cdots\end{bmatrix}$. Let $\mathbf{A}_{ij}$ and $\mathbf{B}_{ij}$ denote the (i,j)th $m_i\times n_j$ and $p_i\times q_j$ blocks of partitioned matrices $\mathbf{A}$ and $\mathbf{B}$, respectively. The Khatri-Rao product for partitioned matrices [45] is defined as $\mathbf{A}\boxplus\mathbf{B} \triangleq \begin{bmatrix} \mathbf{A}_{11}\otimes\mathbf{B}_{11} & \mathbf{A}_{12}\otimes\mathbf{B}_{12} & \cdots \\ \mathbf{A}_{21}\otimes\mathbf{B}_{21} & \mathbf{A}_{22}\otimes\mathbf{B}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$, where $\mathbf{A}_{ij}\otimes\mathbf{B}_{ij}$ and $\mathbf{A}\boxplus\mathbf{B}$ are $m_ip_i\times n_jq_j$ and $(\sum m_ip_i)\times(\sum n_jq_j)$, respectively.

B. Outline

In Section II, we present the JISA framework. We then focus on a special case that is based on SOS, and its algebraic restatement as CBD. We show how identifiability of the former is related to uniqueness of the latter. In Section III, we introduce reducibility and irreducibility, key concepts in our analysis. In Section IV, we review previous related identifiability results. After presenting the required background material, we turn to our main result. Section V presents our main result, a theorem on the identifiability of SOS-based JISA, and accordingly, on the uniqueness of CBD, under certain constraints. In Section VI, we discuss the meaning of this theorem. In Section VII, we illustrate and explain our main results through theoretical examples. In Section VIII, we discuss our results in a broader context, including their relation to ISA, JBD, and tensor decompositions. Section IX concludes our work.

II. Problem Formulation

A. Joint Independent Subspace Analysis (JISA)

Consider an ensemble of $K \geq 2$ datasets, modeled as

$$\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]} \quad , \quad k = 1, \ldots, K \,, \tag{1}$$

where vector $\mathbf{s}^{[k]}$ is an instance of a multivariate random (stochastic) process, and the matrices $\mathbf{A}^{[k]}$ are deterministic and different from each other. Consequently, $\mathbf{x}^{[k]}$, which is a vector of length $I^{[k]} \geq 2 \; \forall k$, is an instance of a multivariate ($I^{[k]}$-variate) random process. In the context of BSS, $\mathbf{s}^{[k]}$ represents signals (sometimes referred to as “sources”), $\mathbf{x}^{[k]}$ represents observations or measurements at $I^{[k]}$ sensors, and $\mathbf{A}^{[k]}$ is a “mixing matrix” representing channel effects between signals and sensors. Therefore, each dataset in (1) is sometimes referred to as a “mixture” of the latent, unobserved, signals.

Each dataset in (1) can always be reformulated as a sum of $R \geq 2$ terms [5]

$$\mathbf{x}^{[k]} = \sum_{i=1}^{R} \mathbf{A}_i^{[k]}\mathbf{s}_i^{[k]} = \sum_{i=1}^{R} \mathbf{x}_i^{[k]} \tag{2}$$

where the ith vector $\mathbf{x}_i^{[k]}$, of length $I^{[k]}$, is modeled as

$$\mathbf{x}_i^{[k]} = \mathbf{A}_i^{[k]}\mathbf{s}_i^{[k]} \,. \tag{3}$$

In this model, $\mathbf{A}^{[k]} = [\mathbf{A}_1^{[k]}| \cdots |\mathbf{A}_R^{[k]}]$ is partitioned column-wise into R blocks $\mathbf{A}_i^{[k]}$ of size $I^{[k]} \times m_i^{[k]}$, where $m_i^{[k]} \geq 1$ $\forall i, k$, and $1 \leq \dim(\operatorname{span}(\mathbf{A}_i^{[k]})) \leq m_i^{[k]}$. Accordingly, $\mathbf{s}_i^{[k]}$, of length $m_i^{[k]}$, is the ith segment of $\mathbf{s}^{[k]} = [\mathbf{s}_1^{[k]\top}, \ldots, \mathbf{s}_R^{[k]\top}]^\top$. This model is more general than classical BSS, in which $K = 1$, $m_i^{[k]} = 1 \; \forall k$, and $\dim(\operatorname{span}(\mathbf{A}_i^{[k]})) = 1 \; \forall i, k$. The K vectors $\mathbf{x}^{[k]}$ can be concatenated in a single vector,

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}^{[1]} \\ \vdots \\ \mathbf{x}^{[K]} \end{bmatrix} = \sum_{i=1}^{R} \begin{bmatrix} \mathbf{A}_i^{[1]} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_i^{[K]} \end{bmatrix} \begin{bmatrix} \mathbf{s}_i^{[1]} \\ \vdots \\ \mathbf{s}_i^{[K]} \end{bmatrix}$$
$$= \sum_{i=1}^{R} (\mathbf{I}_K \boxdot \mathbf{A}_i)\mathbf{s}_i = \sum_{i=1}^{R} \mathbf{x}_i \,, \tag{4}$$

where $\mathbf{x}_i = [\mathbf{x}_i^{[1]\top}, \ldots, \mathbf{x}_i^{[K]\top}]^\top$ is of the same length as $\mathbf{x}$; $\mathbf{s}_i = [\mathbf{s}_i^{[1]\top}, \ldots, \mathbf{s}_i^{[K]\top}]^\top$, and $\mathbf{A}_i = [\mathbf{A}_i^{[1]}| \cdots |\mathbf{A}_i^{[K]}]$. Table I summarizes these notations. This model, which is quite general, will be simplified in Section II-D. The last column of Table I refers to the simplified model. For now, we focus on the first two columns of Table I. The JISA model that we consider in this paper satisfies the following assumptions:

(A1) For a specific i, some or all of the entries of $\mathbf{s}_i$ may be statistically dependent. In particular, entries of $\mathbf{s}_i^{[k]}$ and $\mathbf{s}_i^{[l]}$ may be dependent, for $k \neq l$. The dependence and independence relations among the elements of $\mathbf{s}_i$ may be different from those within $\mathbf{s}_j$, for $i \neq j$.

(A2) $\mathbf{s}_1, \ldots, \mathbf{s}_R$ are statistically independent, with R maximal.

Figure 1 illustrates the model that we have just described.

The JISA framework subsumes several well-known models in the literature. When $K = 1$, this model amounts to ISA [4], [5], [32]. The special case of ISA with $m_i^{[k]} = 1$ $\forall i$ is commonly known as ICA [7]. When $m_i^{[k]} = 1 \; \forall i$ and $K \geq 2$, this model amounts to IVA [6]. Therefore, JISA can be regarded as a model that consists of several ISA problems, linked by statistical dependencies among the latent signals.

Quantity	General	Simplified
$\mathbf{m}^{[k]} = [m_1^{[k]}, \ldots, m_R^{[k]}]^\top$	✓	
$\mathbf{m}_i = [m_i^{[1]}, \ldots, m_i^{[K]}]^\top$	✓	
$M = \sum_{i=1}^{R} m_i$		✓
	Size	
$\mathbf{s}_i^{[k]}$	$m_i^{[k]}$	m_i
$\mathbf{s}^{[k]} = [\mathbf{s}_1^{[k]\top}, \ldots, \mathbf{s}_R^{[k]\top}]^\top$	$\sum_{i=1}^{R} m_i^{[k]} = \mathbf{m}^{[k]\top}\mathbf{1}_R$	M
$\mathbf{s}_i = [\mathbf{s}_i^{[1]\top}, \ldots, \mathbf{s}_i^{[K]\top}]^\top$	$\sum_{k=1}^{K} m_i^{[k]} = \mathbf{m}_i^\top \mathbf{1}_K$	KM
$\mathbf{x}^{[k]}$, $\mathbf{x}_i^{[k]}$	$I^{[k]}$	M
$\mathbf{x}_i = [\mathbf{x}_i^{[1]\top}, \ldots, \mathbf{x}_i^{[K]\top}]^\top$	$\sum_{k=1}^{K} I^{[k]}$	KM
$\mathbf{A}_i^{[k]}$	$I^{[k]} \times m_i^{[k]}$	$M \times m_i$
$\mathbf{A}^{[k]} = [\mathbf{A}_1^{[k]}| \cdots |\mathbf{A}_R^{[k]}]$	$I^{[k]} \times \mathbf{m}^{[k]\top}\mathbf{1}_R$	$M \times M$

TABLE I

Glossary of principal JISA notations. The second column refers to the general model in Section II-A. The third column refers to the simplified model in Section II-D.

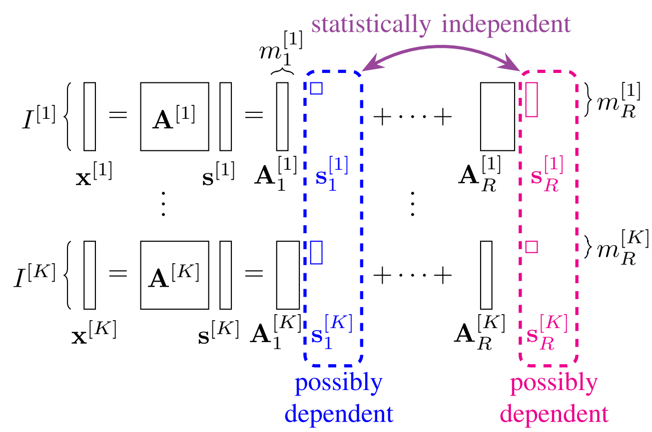


Fig. 1. Illustration of a JISA model with K mixtures and R terms (summands) in each mixture. Random vectors $\mathbf{s}_i^{[k]}$ may have different length, denoted by $m_i^{[k]} \geq 1$. Accordingly, $\mathbf{A}_i^{[k]}$ may have different width. Entries of random vectors $\mathbf{s}_i^{[k]}$ with the same subscript i are allowed to be statistically dependent (but do not have to). Random vectors with different subscripts are always statistically independent.

Assumption (A1) enables the link among datasets by allowing statistical dependence among signals that belong to different datasets. Assumption (A1) enables the multivariate nature of the signals within each dataset by allowing statistical dependence among the $m_i^{[k]} \geq 1$ entries of $\mathbf{s}_i^{[k]}$. Assumption (A2) implies that $\mathbf{s}_i^{[k]}$, the ith signal in dataset k, is always statistically independent of $\mathbf{s}_j^{[l]}$, the jth signal in dataset l, for $i \neq j$ and any (k, l). Together with (4), assumption (A2) implies that $\mathbf{x}_1, \ldots, \mathbf{x}_R$ are statistically independent as well. Therefore, the requirement for maximal R in assumption (A2) implies that JISA can separate $\mathbf{x}$ into a sum of at most R statistically independent random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_R$. Whereas assumption (A1) allows each $m_i^{[k]}$ to be larger than one, assumption (A2) has the opposite effect, of making sure that each $m_i^{[k]}$ has the smallest possible value within the JISA framework.

We emphasize that the values of $m_i^{[k]}$, R, and the respective partition of $\mathbf{A}^{[k]}$ and $\mathbf{s}^{[k]}$ that we use in this paper for our JISA analysis, are w.r.t. the statistical properties of the overall JISA model, and not w.r.t. each underlying ISA mixture individually. The simplest example is a mixture of Gaussian multivariate i.i.d. signals, whose covariance matrix can always be diagonalized, which implies that effectively, all signals are one-dimensional. However, as soon as we consider three or more such mixtures in a single JISA framework, it is possible to define signals of dimensions larger than one within each mixture. A large part of this paper is dedicated to the mathematical formulation and understanding of this phenomenon. This also means that in general, the JISA model cannot provide a finer separation of the observations beyond these R statistically independent terms.

It is clear from (3) that one cannot distinguish between the pairs $(\mathbf{A}_i^{[k]}, \mathbf{s}_i^{[k]})$ and $(\mathbf{A}_i^{[k]}\mathbf{Z}_{ii}^{-[k]}, \mathbf{Z}_{ii}^{[k]}\mathbf{s}_i^{[k]})$, where $\mathbf{Z}_{ii}^{[k]}$ is an arbitrary nonsingular $m_i^{[k]} \times m_i^{[k]}$ matrix. This means that only $\mathbf{x}_i^{[k]}$ and $\operatorname{span}(\mathbf{A}_i^{[k]})$, which do not suffer from this inherent unavoidable ambiguity, may be uniquely identified using the JISA framework (unless additional assumptions are imposed, which is not the case in this paper). Therefore, within each mixture, the inherent ambiguities of JISA are the same as those of an ISA problem that has the same partition into R statistically independent elements. For a clearer view of how these inherent ambiguities are manifested in the joint framework, let us look at (4). It follows from (4) that the pair $(\mathbf{I}_K \boxplus \mathbf{A}_i\,, \mathbf{s}_i)$ is indistinguishable from $\big((\mathbf{I}_K \boxplus \mathbf{A}_i)(\oplus_{k=1}^K \mathbf{Z}_{ii}^{-[k]})\,, (\oplus_{k=1}^K \mathbf{Z}_{ii}^{[k]})\mathbf{s}_i\big)$, because the product of the terms in each pair is $\mathbf{x}_i$. Therefore, only $\mathbf{x}_i$ and the subspace associated with the ith signal, $\operatorname{span}(\mathbf{I}_K \boxplus \mathbf{A}_i)$, which do not suffer from this inherent unavoidable ambiguity, may be uniquely identified using the JISA framework. Hence, JISA can be regarded as a (joint) subspace estimation problem.

We thus define the problem associated with JISA as follows: given $\mathbf{x}$, and given $\{m_i^{[k]}\}_{i=1,k=1}^{R\;,\;K}$ [1], obtain statistically independent $\mathbf{x}_1, \ldots, \mathbf{x}_R$. In practice, the distributions are estimated from data, and therefore, this goal can be achieved only approximately. Accordingly, we suggest the following definition of JISA uniqueness and identifiability:

Definition II.1. If, for a given $\mathbf{x}$, and given $\{m_i^{[k]}\}_{i=1,k=1}^{R\;,\;K}$, any choice of $\mathbf{x}_1, \ldots, \mathbf{x}_R$ that satisfy all our model assumptions yields the same R summands in (4), we say that the factorization of $\mathbf{x}$ into a sum of R terms is unique, and that the JISA model is identifiable.

By Definition II.1, non-identifiability means that for the same observations, there exists another set of random vectors $\{\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_R\}$ that sum up to $\mathbf{x}$, but with $\widehat{\mathbf{x}}_i \neq \mathbf{x}_i$ for at least one value of i (obviously, to balance the equation, this must hold for at least two values of i). These $\widehat{\mathbf{x}}_1, \ldots, \widehat{\mathbf{x}}_R$ are associated with mixing matrices that we denote $\widehat{\mathbf{A}}^{[1]}, \ldots, \widehat{\mathbf{A}}^{[K]}$. This means that $\operatorname{span}(\mathbf{A}_i^{[k]}) \neq \operatorname{span}(\widehat{\mathbf{A}}_i^{[k]})$ and $\operatorname{span}(\mathbf{A}_j^{[k]}) \neq \operatorname{span}(\widehat{\mathbf{A}}_j^{[k]})$ for at least one value of k and one pair of $i \neq j$, that is, not all signal subspaces have been properly identified.

[1] In certain cases, it may be possible to estimate $\{m_i^{[k]}\}$ directly from the data; however, such methods are beyond the scope of this work.

The aim of this paper is to provide the necessary and sufficient conditions that guarantee that this separation be unique, under certain additional assumptions. Later on in this section, and in Section III, we shall give a more concrete meaning to these assumptions, in terms of SOS.

B. Second-Order Statistics (SOS)

In this paper, we focus on SOS. Therefore, in this section, we show how assumptions (A1) and (A2) are manifested in the SOS. The cross-correlation between any two random vectors $\mathbf{s}_i^{[k]}$ and $\mathbf{s}_j^{[l]}$ satisfies

$$\mathbf{S}_{ij}^{[k,l]} \triangleq \mathrm{E}\{\mathbf{s}_i^{[k]}\mathbf{s}_j^{[l]\mathsf{H}}\} = \begin{cases} \mathbf{S}_{ii}^{[k,l]} & i = j \\ \mathbf{0} & i \neq j \end{cases} . \tag{5}$$

The zero values on the right-hand side (RHS) of (5) are due to assumption (A2). Assumption (A1) implies that $\mathbf{S}_{ii}^{[k,l]}$ may be non-zero. The $m_i^{[k]} \times m_j^{[l]}$ matrix $\mathbf{S}_{ij}^{[k,l]}$ can be placed in the (k,l)th block of the $\mathbf{m}_i^\top \mathbf{1}_K \times \mathbf{m}_j^\top \mathbf{1}_K$ matrix

$$\mathbf{S}_{ij} \triangleq \mathrm{E}\{\mathbf{s}_i\mathbf{s}_j^{\mathsf{H}}\} = \begin{bmatrix} \mathbf{S}_{ij}^{[1,1]} & \cdots & \mathbf{S}_{ij}^{[1,K]} \\ \vdots & & \vdots \\ \mathbf{S}_{ij}^{[K,1]} & \cdots & \mathbf{S}_{ij}^{[K,K]} \end{bmatrix} = \begin{cases} \mathbf{S}_{ii} & i = j \\ \mathbf{0} & i \neq j \end{cases} \tag{6}$$

as well as in the (i,j)th block of the $\mathbf{m}^{[k]\top}\mathbf{1}_R \times \mathbf{m}^{[l]\top}\mathbf{1}_R$ matrix

$$\mathbf{S}^{[k,l]} \triangleq \mathrm{E}\{\mathbf{s}^{[k]}\mathbf{s}^{[l]\mathsf{H}}\} = \mathbf{S}_{11}^{[k,l]} \oplus \cdots \oplus \mathbf{S}_{RR}^{[k,l]} . \tag{7}$$

The block-diagonal structure of $\mathbf{S}^{[k,l]}$ follows from (5). The RHS of Fig. 2(a) illustrates (7). Figure 2(b) illustrates (5) and (6).

C. Coupled Block Diagonalization (CBD)

The cross-correlation between observations in any two datasets k and l satisfies

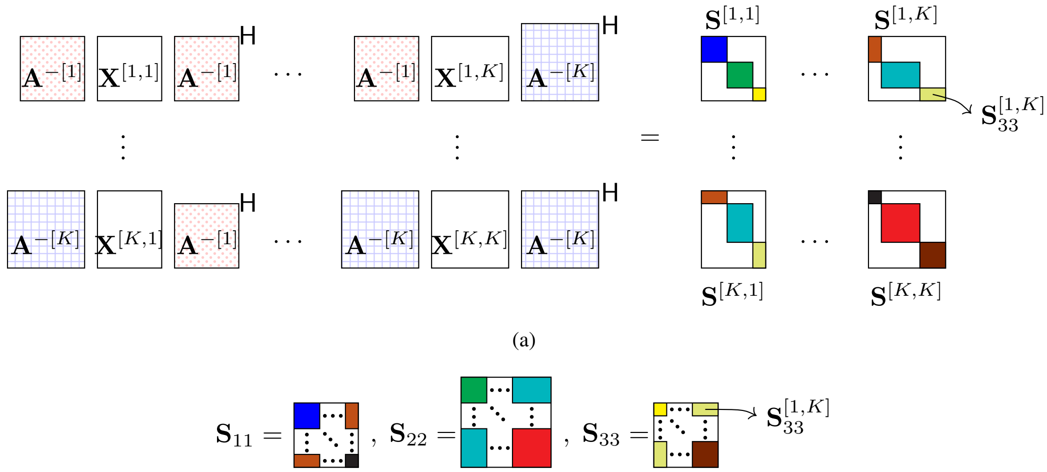
$$\mathbf{X}^{[k,l]} \triangleq \mathrm{E}\{\mathbf{x}^{[k]}\mathbf{x}^{[l]\mathsf{H}}\} = \mathbf{A}^{[k]}\mathbf{S}^{[k,l]}\mathbf{A}^{[l]\mathsf{H}} \tag{8a}$$

$$= \sum_{i=1}^{R} \mathbf{A}_i^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{A}_i^{[l]\mathsf{H}} \tag{8b}$$

where the RHS of (8a) is due to (1), and (8b) is due to (3) and (7). Due to the block partition of $\mathbf{A}^{[k]}$, we refer to the decomposition of each $\mathbf{X}^{[k,l]}$ in (8) as a “block decomposition”, and to the decomposition all at once of the ensemble $\{\mathbf{X}^{[k,l]}\}_{k,l=1}^K$ as a *coupled* block decomposition. If $\mathbf{A}^{[k]}$ is nonsingular $\forall k$, (8) can be rewritten as

$$\mathbf{A}^{-[k]}\mathbf{X}^{[k,l]}\mathbf{A}^{-[l]\mathsf{H}} = \mathbf{S}^{[k,l]} \tag{9}$$

where the RHS is block-diagonal by (7). For fixed (k,l), (9) represents a block-diagonalization of $\mathbf{X}^{[k,l]}$ by two transformation matrices, $\mathbf{A}^{-[k]}$ and $\mathbf{A}^{-[l]}$. Therefore, when applied to all k,l at once, (9) amounts to *coupled block diagonalization (CBD)* [11]. CBD is illustrated in Fig. 2(a). Similarly, in analogy to the more familiar notion of matrix congruence, we use the term *coupled congruence* to denote the relation in (9) between $\{\mathbf{X}^{[k,l]}\}_{k,l=1}^K$ and $\{\mathbf{S}^{[k,l]}\}_{k,l=1}^K$ via the K nonsingular



(a)

(b)

Fig. 2. Example of SOS-based JISA with $K \geq 2$ real-valued datasets, each with $R = 3$ independent elements of rank $m_1^{[1]} = m_2^{[1]} = m_3^{[K]} = 2$, $m_1^{[K]} = m_3^{[1]} = 1$, $m_2^{[K]} = 3$, and nonsingular (un)mixing matrices of size $I^{[1]} \in \mathbb{R}^{5\times 5}$ and $I^{[K]} \in \mathbb{R}^{6\times 6}$. (a) SOS-based JISA by CBD of $\{\mathbf{X}^{[k,l]}\}_{k,l=1}^{K}$. Note the block-diagonal structure of $\mathbf{S}^{[k,l]}$ and the symmetry along the main axis $k = l$. (b) Rearranging the elements of $\{\mathbf{S}^{[k,l]}\}_{k,l=1}^{K}$ in $\{\mathbf{S}_{ii}\}_{i=1}^{R}$.

transformation matrices $\mathbf{A}^{[k]}$. One can readily verify that coupled congruence is an equivalence relation. Equivalence via coupled congruence is essential for JISA identifiability and CBD uniqueness, as we shall see later on.

In accordance with our discussion of the inherent JISA indeterminacies earlier in Section II-A, each summand in (8b) satisfies

$$\mathbf{A}_i^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{A}_i^{[l]\mathsf{H}} = (\mathbf{A}_i^{[k]}\mathbf{Z}_{ii}^{-[k]})(\mathbf{Z}_{ii}^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{Z}_{ii}^{[l]\mathsf{H}})(\mathbf{Z}_{ii}^{-[l]\mathsf{H}}\mathbf{A}_i^{[l]\mathsf{H}})\,. \tag{10}$$

In analogy to Definition II.1, we can now suggest a definition for the uniqueness of CBD:

Definition II.2. If, for given $\{\mathbf{X}^{[k,l]}\}_{k,l=1}^{K}$, and given $\{m_i^{[k]}\}_{i=1,k=1}^{R\ ,\ K}$, any choice of $\{\mathbf{A}_i^{[k]}\}_{i=1,k=1}^{R\ ,\ K}$ and $\{\mathbf{S}_{ii}^{[k,l]}\}_{k,l=1}^{K}$ that satisfy (8) yields the same R summands $\mathbf{A}_i^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{A}_i^{[l]\mathsf{H}}$ $\forall k,l$, we say that the decomposition in (8) (and (9), if applicable) is essentially unique[2].

Alternatively, we can restate Definition II.2 by replacing the R summands with the R $\mathbf{m}_i^\top\mathbf{1}_K$-dimensional signal subpaces $\operatorname{span}(\mathbf{I}_K \boxminus \mathbf{A}_i)$. Example VII.1 in Section VII illustrates Definition II.2.

D. JISA via CBD and Additional Model Assumptions

In this paper, we focus on a JISA model in which, in addition to assumptions (A1) and (A2), we assume that

(A3) Each underlying multivariate random process is real-valued, normally distributed: $\mathbf{s}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{S}_{ii})$, with i.i.d. samples.
(A4) $\mathbf{A}^{[k]}$ is nonsingular and real-valued $\forall k$,
(A5) $\mathbf{S}_{ii}$ is nonsingular $\forall i$.
(A6) Each $\mathbf{S}_{ii}^{[k,l]}$ with $k \neq l$ is either zero-valued or full-rank.
(A7) $m_i^{[k]} = m_i$ $\forall i, k$.

[2]The term "essentially unique" emphasizes the fact that there still exists an inherent unavoidable ambiguity, as defined by (10).

Assumptions (A3), (A4) and (A7) imply that $\mathbf{s}_i^{[k]} \in \mathbb{R}^{m_i\times 1}$, $\mathbf{s}^{[k]} \in \mathbb{R}^{M\times 1}$, $\mathbf{s}_i \in \mathbb{R}^{Km_i\times 1}$, $\mathbf{A}_i^{[k]} \in \mathbb{R}^{M\times m_i}$, $\mathbf{A}^{[k]} \in \mathbb{R}^{M\times M}$, $\mathbf{x}^{[k]}$ and $\mathbf{x}_i^{[k]} \in \mathbb{R}^{M\times 1}$, $\mathbf{x}$ and $\mathbf{x}_i \in \mathbb{R}^{KM\times 1}$, where $M \triangleq \sum_{i=1}^{R} m_i$. These quantities are summarized in the last column of Table I.

When assumption (A3) holds, the correlation matrices $\{\mathbf{X}^{[k,l]}\}_{k,l=1}^{K}$ are sufficient statistics for full model identifiability [1][3]. In this case, JISA amounts to coupled block decomposition (or to CBD, if assumption (A4) holds as well) of $\{\mathbf{X}^{[k,l]}\}_{k,l=1}^{K}$, as formulated in Section II-C. In this case, Definition II.2 now establishes the link between the uniqueness of CBD and the identifiability of JISA. When assumption (A3) holds, each mixture in (1) is, in general, not identifiable individually, as mentioned in Section I. However, previous results (see Section IV) provide supporting evidence to the identifiability of the joint decomposition when $K \geq 2$. Hence, this JISA model highlights the added value of JISA w.r.t. analysing each mixture individually. When assumption (A4) holds, our model's identifiability does not depend on $\mathbf{A}^{[k]}$. A useful implication of assumptions (A3) to (A5) is that the FIM can be derived in closed form [1] (see also Footnote 3). The proof of our main results, in Section V, relies on characterizing all the cases in which the FIM is singular. In Appendix B, we show that characterizing the singularity of the FIM boils down to characterizing the non-trivial solutions to a system of coupled matrix equations. The coefficients of these matrix equations are the source correlations. Assumptions (A6) and (A7) impose constraints on the coefficients of the coupled matrix equations that we have just mentioned, and thus, further simplify our derivations. In this paper, we shall fully characterize the necessary and sufficient additional constraints on $\mathbf{S}_{ii}$ that guarantee identifiability, and specify those values of $\mathbf{S}_{ii}$ for which the model is not identifiable, given these assumptions.

[3]In [1] we assumed, for clarity of exposition, that $m_i^{[k]} = m_i$ $\forall k$. The generalization of all the results in [1] to $m_i^{[k]} \neq m_i^{[l]}$ for every $l \neq k$ is trivial and straightforward, up to the minimal necessary notational adaptations.

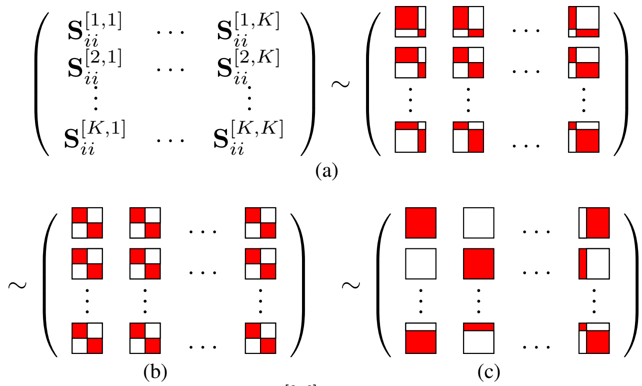


Fig. 3. A block matrix $\mathbf{S}_{ii} = \{\mathbf{S}_{ii}^{[k,l]}\}$ reducible by coupled congruence. "$\sim$" denotes equivalence by coupled congruence. White: zero, color: any value. In this example, $\mathbf{S}_{ii}^{[k,l]} \in \mathbb{R}^{m_i \times m_i}$ with $m_i = 4$, and the size of the zero sub-blocks on the RHS is determined, according to (11), by (a) $\alpha^{[1]} = 1$, $\alpha^{[2]} = 2$, and $\alpha^{[K]} = 3$, (b) $\alpha^{[1]} = \alpha^{[2]} = \alpha^{[K]} = 2$, (c) $\alpha^{[1]} = 4$, $\alpha^{[2]} = 0$, and $\alpha^{[K]} = 3$.

III. Reducibility and Irreducibility

The concept of (ir)reducibility is fundamental to our analysis. We begin with definitions, and then explain how they are related to our model assumptions.

Definition III.1. A block matrix $\mathbf{S}_{ii}$ whose (k,l)th block is $\mathbf{S}_{ii}^{[k,l]}$, $k,l = 1,\ldots,K$, is said to be *reducible by coupled congruence* if there exist K nonsingular matrices (transformations) $\mathbf{T}_{ii}^{[k]}$ such that

$$\mathbf{T}_{ii}^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{T}_{ii}^{[l]\top} = \begin{bmatrix} \mathbf{S}_{i_1 i_1}^{[k,l]} & \mathbf{0} \\ \mathbf{0}_{\alpha^{[k]}\times\beta^{[l]}} & \mathbf{S}_{i_2 i_2}^{[k,l]} \end{bmatrix} \quad \forall k,l\,, \qquad (11)$$

where $\alpha^{[k]}$ and $\beta^{[k]} \triangleq m_i - \alpha^{[k]}$ are nonnegative integers $\forall k$, and positive for at least one k. Otherwise, $\mathbf{S}_{ii}$ is said to be *irreducible by coupled congruence*.

Examples of a block matrix $\mathbf{S}_{ii}$ that is reducible by coupled congruence are illustrated in Fig. 3. Figures 3(a) and 3(b) illustrate two cases where $\alpha^{[k]}$ and $\beta^{[k]}$ are positive $\forall k$. In Fig. 3(c), one $\alpha^{[k]}$ and one $\beta^{[k']}$, $k \neq k'$, are zero. The zero blocks in Fig. 3(c) indicate that the corresponding blocks of $\mathbf{S}_{ii}$ were originally zero. Note that all the off-diagonal blocks in Figs. 3(a) and 3(c) are singular matrices, which means that the corresponding blocks in $\mathbf{S}_{ii}$ were singular as well. In Fig. 3(b), all the blocks are nonsingular matrices, which means that all the blocks of $\mathbf{S}_{ii}$ were nonsingular as well.

Definition III.1 can be restated as saying that $\mathbf{S}_{ii}$ is reducible by coupled congruence if and only if (iff) it is equivalent by coupled congruence to another matrix, whose (k,l)th block has a block-diagonal structure as in (11), $\forall k,l$. Definition III.1 implies that $\mathbf{S}_{ii}$ cannot be reducible by coupled congruence if $m_i = 1$. If the transformations in Definition III.1 are all unitary, we say that $\mathbf{S}_{ii}$ is *unitarily (ir)reducible* in the coupled sense. Clearly, if a matrix is irreducible by coupled congruence, it is irreducible by coupled unitary transformations. The converse, however, does not generally hold, because a matrix may be irreducible by coupled unitary transformations yet reducible by non-unitary coupled congruence transformations. As an example, consider the case $K = 2$ with $m_i \geq 2$, which is always reducible by generalized eigenvalue decomposition (GEVD), as explained in Section IV; however, the reducing transformations are not unitary, except for very special cases. Within the JISA framework, however, we can take advantage of the fact that the coupled congruence via any arbitrary transformation matrices $\mathbf{S}_{ii}^{[k,l]} \leftrightarrow \mathbf{Z}_{ii}^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{Z}_{ii}^{[l]\top}$ $\forall k,l$ in (10) can be regarded as basis transformations on the rows and columns of the blocks of $\mathbf{S}_{ii}$ (in analogy to [46, Remark 7]). For brevity, we refer to this operation as *coupled basis transformation*. We can now state the following:

Proposition III.2. $\mathbf{S}_{ii}$ is reducible by coupled congruence iff there exists a coupled basis transformation in which the transformed $\mathbf{S}_{ii}$ is reducible by coupled unitary transformations.

Proof. Let $\{\mathbf{Z}_{ii}^{[k]}\}_{k=1}^{K}$ be non-unitary transformations such that $\mathbf{S}_{ii}$ is reducible by coupled congruence (Definition III.1). Let $\mathbf{Z}_{ii}^{[k]} = \mathbf{Q}^{[k]}\mathbf{R}^{[k]}$ denote the QR decomposition of $\mathbf{Z}_{ii}^{[k]}$, with $\mathbf{Q}^{[k]}$ unitary and $\mathbf{R}^{[k]}$ upper triangular $\forall k$. Then,

$$\underbrace{\mathbf{Q}^{[k]}\mathbf{R}^{[k]}}_{\text{non-unitary}}\mathbf{S}_{ii}^{[k,l]}\mathbf{R}^{[l]\top}\mathbf{Q}^{[l]\top} = \begin{bmatrix} \mathbf{S}_{i_1 i_1}^{[k,l]} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{i_2 i_2}^{[k,l]} \end{bmatrix} \quad \forall k,l \qquad (12)$$

which implies that $\{\mathbf{R}^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{R}^{[l]\top}\}_{k,l=1}^{K}$ is unitarily reducible by coupled congruence. $\square$

Proposition III.2 amounts to saying that $\mathbf{S}_{ii}$ is irreducible by coupled congruence iff there does not exist any coupled basis transformation in which the transformed $\mathbf{S}_{ii}$ is reducible by a coupled unitary transformation. The concept of irreducibility in general, and Proposition III.2 in particular, is a key ingredient for proving our main result, Theorems V.1 and V.2.

We now turn to explaining the relation between reducibility and our model. Assumptions (A6) and (A7) imply that a full-rank $\mathbf{S}_{ii}^{[k,l]}$ is square and nonsingular. This excludes all the reducible cases in which at least one $\alpha^{[k]}$ or $\beta^{[k]}$ are zero, because these cases correspond to patterns in which one or more $\mathbf{S}_{ii}^{[k,l]}$ have a row (or column) of zeros and are thus rank deficient, as in Fig. 3(c). Assumptions (A6) and (A7) exclude also reducible cases with $\alpha^{[k]}$ positive $\forall k$ and $\alpha^{[k]} \neq \beta^{[k]}$ for at least one k, because they are associated with blocks that are a direct sum of rectangular non-square matrices, and thus singular, as in Fig. 3(a). Cases with positive $\alpha^{[k]}$ and $\beta^{[k]}$ $\forall k$, as in Figs. 3(a) and 3(b) are eliminated by assumption (A2) (indeed, there is an overlap: cases as in Fig. 3(a) are excluded by both arguments). To see this, suppose for a moment that all our model assumptions hold, except assumption (A2), which is violated such that R is not maximal. Then, there exists a random vector $\mathbf{s}_i$ with $m_i \geq 2$ that can be written as two distinct statistically independent random vectors that we denote $\mathbf{s}_{i_1}$ and $\mathbf{s}_{i_2}$. Then, in each dataset, similarly to our discussion in Section II-A, we can write $\mathbf{s}_i^{[k]} = [\mathbf{s}_{i_1}^{[k]\top}\ \mathbf{s}_{i_2}^{[k]\top}]^\top$, where $\mathbf{s}_{i_1}^{[k]}$ and $\mathbf{s}_{i_2}^{[k]}$ have length $m_{i_1}^{[k]}$ and $m_{i_2}^{[k]}$, respectively, $m_{i_1}^{[k]} + m_{i_2}^{[k]} = m_i$. Accordingly, $\mathbf{s}_{i_1}$ and $\mathbf{s}_{i_2}$ have length $\sum_{k=1}^{K} m_{i_1}^{[k]}$ and $\sum_{k=1}^{K} m_{i_2}^{[k]}$, respectively. Our assumption of statistical independence (which has now boiled down to decorrelation, due to assumption (A3)) among $\mathbf{s}_{i_1}$ and $\mathbf{s}_{i_2}$ implies that $\mathrm{E}\{\mathbf{s}_{i_1}^{[k]}\mathbf{s}_{i_2}^{[l]\top}\} = \mathbf{0}_{m_{i_1}^{[k]} \times m_{i_2}^{[l]}}$ $\forall k, l$. Hence,

$$\mathbf{S}_{ii}^{[k,l]} = \mathrm{E}\{\mathbf{s}_i^{[k]}\mathbf{s}_i^{[l]\top}\} = \begin{bmatrix} \mathbf{S}_{i_1 i_1}^{[k,l]} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{i_2 i_2}^{[k,l]} \end{bmatrix} \quad \forall k, l \,. \tag{13}$$

which is reducible by Definition III.1. We conclude that the correlation matrices $\mathbf{S}_{ii}$ of a JISA model satisfying assumptions (A1) to (A7) are irreducible by coupled congruence.

IV. Previous Related Identifiability Results

In this section, we briefly review previous related results.

One can readily verify that if only one dataset is considered (i.e., $K = 1$), the SOS of the observations (8) do not provide a sufficient number of constraints w.r.t. the free model parameters (e.g., [47, Section V.A], [1, Section V.A]). Therefore, a single dataset in (1) is not identifiable given our model assumptions. This non-identifiability is a multivariate generalization to the well-known fact that Gaussian processes with i.i.d. samples cannot be separated blindly (e.g., [8]). For this reason, a large body of literature has dealt with achieving uniqueness to this single-set scenario by exploiting various types of diversity, including non-stationarity, dependence among samples, HOS, complex-valued data, and additional observational modes. Most of this literature deals with the $m_i = 1$ case; see, e.g, [9], [13], for a comprehensive review. The $m_i \geq 2$ case has received much less attention in the literature; we refer, e.g., to [4], [5], [19], [32], [46] and references therein, for some pertinent results, in the single-set context. In this paper, however, we focus on a different approach, of showing that a multiset framework may be sufficient for identifiability, even if none of the "classical" types of diversity, which guarantee uniqueness and identifiability in a single-set setup, are present. In other words, we use a link between different mixtures as an alternative form of diversity [13]. In order to clarify the relationship between this new type of diversity and previous single-set results, we demonstrate in Section VIII a link between the identifiability of piecewise stationary ISA and the JISA identifiability results in this paper.

For $m_i^{[k]} = 1$ $\forall i, k$, it has been argued in [48], and later rigorously proven in [14], [17], [18], that as soon as $K \geq 2$, overdetermined IVA of Gaussian processes with i.i.d. samples is identifiable, except very particular cases.

The case of JISA with $K = 2$ datasets deserves special attention, as we now explain. When $K = 2$, the JISA model can be reformulated as a GEVD [49, Chapter 12.2, Equation (53)] (see also [50, Sec. 4.3]). In this case, estimates of $\{\mathbf{A}^{[k]}\}_{k=1}^2$, which we denote $\{\mathbf{A}^{[k]\mathrm{GEVD}}\}_{k=1}^2$, can be obtained in closed form, from the generalized eigenvectors. Given our model assumptions, these estimates always exist. Furthermore, they always achieve exact diagonalization of $\{\mathbf{X}^{[k,l]}\}_{k,l=1}^2$: $\mathbf{A}^{-[k]\mathrm{GEVD}}\mathbf{X}^{[k,l]}\mathbf{A}^{-[l]\mathrm{GEVD}\top}$ is a diagonal matrix $\forall k, l$, regardless of the size of the blocks on the diagonal of $\{\mathbf{S}^{[k,l]}\}_{k,l=1}^2$, i.e., even if the input latent data has $m_i \geq 2$ $\forall k$. This means that from the point of view of the JISA model, if $K = 2$, then $m_i^{[k]} = 1$ $\forall i, k$, and $\mathbf{S}^{[k,l]}$ are exactly diagonal $\forall k, l$. Therefore, in this case, JISA is tantamount to IVA. Accordingly, the identifiability of this special case can be derived directly using algebraic arguments on the uniqueness of the GEVD: the one-dimensional subspaces associated with the columns of $\mathbf{A}^{[k]\mathrm{GEVD}}$ are uniquely identified iff the generalized eigenvalues of the GEVD are distinct. Naturally, this identifiability result coincides with previous results on IVA [14], [17], [18]. In Section VIII-C, we shall explain why the uniqueness of the GEVD provides sufficient, but not necessary, conditions for JISA identifiability.

In [1], we presented supporting evidence, in terms of a balance of degrees of freedom (d.o.f.), as well as numerical experiments that demonstrated the invertibility of the FIM, that SOS-based JISA with $m_i^{[k]} = m_i \geq 1$ $\forall k$ can be identifiable. In this paper, we prove this rigorously, by characterizing all the cases in which this JISA model is not identifiable, and showing that non-identifiability occurs only in very particular cases. We point out that although, in many cases, the same conclusions hold also when $m_i^{[k]} \neq m_i$, this is not always the case, as recently demonstrated in [21]. The full analysis of this more general scenario will be discussed in a different publication.

V. Main Result: JISA Identifiability

The main contribution of this paper is providing the necessary and sufficient conditions for the identifiability of a JISA model satisfying assumptions (A1) to (A7). In Section II-D we explained that given assumptions (A1) to (A7), JISA is tantamount to CBD. Therefore, the following results on JISA identifiability also characterize the uniqueness of CBD, when the matrices in (9) satisfy the constraints corresponding to assumptions (A1) to (A7). Uniqueness and identifiability of these models were defined in Definitions II.1 and II.2 and Section II-D. Further implications of the results of this section are discussed in Sections VIII and IX.

Let us begin with the simpler case, where $\mathbf{S}_{ii}^{[k,l]}$ are nonsingular matrices $\forall i, k, l$.

Theorem V.1. *Consider a JISA model satisfying assumptions (A1) to (A7), and, in addition, $\mathbf{S}_{ii}^{[k,l]}$ are nonsingular matrices $\forall i, k, l$. This JISA model is not identifiable iff there exists at least one pair $(i, j) \in \{1, \ldots, R\}^2$, $i \neq j$, for which $m_j = m_i$ and*

$$\mathbf{S}_{jj}^{[k,l]} = \mathbf{\Psi}^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{\Psi}^{[l]\top} \quad \forall k, l \tag{14}$$

where $\{\mathbf{\Psi}^{[k]}\}_{k=1}^K$ are nonsingular $m_i \times m_i$ matrices.

The proof of Theorem V.1 is given in Appendix C.

Theorem V.1 implies that if (14) holds for some pair (i, j), then $\mathrm{span}(\mathbf{I}_K \boxminus \mathbf{A}_i)$, which is the subspace associated with the ith signal, cannot be distinguished from $\mathrm{span}(\mathbf{I}_K \boxminus \mathbf{A}_j)$, which is the subspace associated with the jth signal. To see how this condition on the signal covariances propagates into the mixing matrices, set (14) in (8b):

$$\mathbf{X}^{[k,l]} = \cdots + \mathbf{A}_i^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{A}_i^{[l]\top} + \cdots + \overbrace{\mathbf{A}_j^{[k]}\mathbf{S}_{jj}^{[k,l]}\mathbf{A}_j^{[l]\top}}^{\mathbf{A}_j^{[k]}\mathbf{\Psi}^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{\Psi}^{[l]\top}\mathbf{A}_j^{[l]\top}} + \cdots \tag{15}$$

Equation (15) implies that the jth term in (8b) can be regarded as having the same covariance matrix $\mathbf{S}_{ii}^{[k,l]}$ as the ith term, $\forall k,l$, but with different mixing matrices associated with it. The non-identifiability arises from the fact that from the point of view of the CBD, the triplets $(\mathbf{A}_j^{[k]}, \mathbf{A}_j^{[l]}, \mathbf{S}_{jj}^{[k,l]})$ and $(\mathbf{A}_j^{[k]}\mathbf{\Psi}^{[k]}, \mathbf{A}_j^{[l]}\mathbf{\Psi}^{[l]}, \mathbf{S}_{ii}^{[k,l]})$ are indistinguishable $\forall k,l$. Example VII.1 in Section VII illustrates a scenario that is non-identifiable by Theorem V.1. Theorem V.1 and (14) will be discussed in detail in Sections VI and VIII.

We now turn to the more general case, where some of $\mathbf{S}_{ii}^{[k,l]}$ may be zero.

Theorem V.2. *Consider a JISA model satisfying assumptions (A1) to (A7). This JISA model is not identifiable iff there exists at least one pair $(i,j) \in \{1,\ldots,R\}^2$, $j \neq i$, for which at least one of the following scenarios holds:*

Scenario 1. Without loss of generality (W.l.o.g.), there exists $\mathcal{D} = \{1,\ldots,D\}$, $2 \leq D \leq K$, D minimal, such that $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ can be written as

$$\mathbf{S}_{ii} = \mathbf{S}_{ii}^{[1:D,1:D]} \oplus \mathbf{S}_{ii}^{[D+1:K,D+1:K]} \tag{16a}$$

$$\mathbf{S}_{jj} = \mathbf{S}_{jj}^{[1:D,1:D]} \oplus \mathbf{S}_{jj}^{[D+1:K,D+1:K]} \tag{16b}$$

where $m_i = m_j$, and there exist D nonsingular matrices $\mathbf{\Psi}^{[k]} \in \mathbb{R}^{m_i \times m_i}$ such that

$$\mathbf{S}_{jj}^{[k,l]} = \mathbf{\Psi}^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{\Psi}^{[l]\top} \quad \textit{for each} \quad k \in \mathcal{D}, l \in \mathcal{D}\,. \tag{17}$$

Scenario 2. W.l.o.g., there exists $\mathcal{D} = \{1,\ldots,D\}$, $D \geq 1$, such that $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ can be written as

$$\mathbf{S}_{ii} = (\oplus_{k=1}^{D}\mathbf{S}_{ii}^{[k,k]}) \oplus \mathbf{S}_{ii}^{[D+1:K,D+1:K]} \tag{18a}$$

$$\mathbf{S}_{jj} = (\oplus_{k=1}^{D}\mathbf{S}_{jj}^{[k,k]}) \oplus \mathbf{S}_{jj}^{[D+1:K,D+1:K]}\,. \tag{18b}$$

The following remarks are in order.

The "w.l.o.g." in Theorem V.2 implies that for each scenario, and for each pair of (i,j), the order the datasets, which is anyway arbitrary, can be modified.

For each pair of (i,j), $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ may satisfy one or more instances of Scenario 1 and/or Scenario 2. For a specific i, several instances of Scenario 1 can hold, but only for disjoint datasets, due to the requirement for minimal D. Several instances of Scenario 2 may overlap. It is possible to add the condition that D be maximal in Scenario 2, to avoid this overlap. The key point is that the model is not identifiable as soon as one of these scenarios holds, for at least one pair of (i,j).

Scenario 1 in Theorem V.2 implies that if (17) holds for some pair (i,j), then $\{\mathrm{span}(\mathbf{A}_i^{[k]})\}_{k=1}^{D}$ cannot be distinguished from $\{\mathrm{span}(\mathbf{A}_j^{[k]})\}_{k=1}^{D}$. The explanation is essentially the same as for Theorem V.1; note that Theorem V.1 is a special case of Theorem V.2, when $D = K$ and there are no zero blocks in any $\mathbf{S}_{ii}$.

The special structures in Scenario 1and in Scenario 2 impose constraints on the dimensions of the signal subspaces, as the following proposition suggests:

Proposition V.3. If, for fixed i, we can write $\mathbf{S}_{ii}$, w.l.o.g., in the direct sum form (16a) with $D = 1$ or $D = 2$, then, for this specific i, $m_i = 1$, that is, $\mathbf{S}_{ii}^{[k,l]}$ are scalars $\forall k,l$.

Proof. Let $\mathbf{S}_{ii}$ satisfy (16a) with $D < K$. It follows from (10) (see also Section III) that $\mathbf{S}_{ii}^{[1:D,1:D]}$ is subject to the coupled congruence transformation $\mathbf{Z}_{ii}^{[k]}\mathbf{S}_{ii}^{[k,l]}\mathbf{Z}_{ii}^{[l]\top}$, where $k,l \in \mathcal{D}$, and $\mathbf{Z}_{ii}^{[k]}$ are artbirary nonsingular matrices. In general, for arbitrary values of $\mathbf{S}_{ii}^{[k,l]}$, when $D \geq 3$, there do not exist $\{\mathbf{Z}_{ii}^{[k]}\}_{k=1}^{D}$ that exactly diagonalize, via coupled congruence, all the blocks of $\mathbf{S}_{ii}^{[1:D,1:D]}$. However, for $D \leq 2$, such transformations always exist, as we now explain. When $D = 1$, $\mathbf{Z}_{ii}^{[1]}$ can always be chosen to diagonalize $\mathbf{S}_{ii}^{[1,1]}$, for example, by singular value decomposition (SVD). For $D = 2$, this exact diagonalization always exists by GEVD, as explained in Section IV. Either way, the fact that the blocks of $\mathbf{S}_{ii}^{[1:D,1:D]}$ can be exactly diagonalized within the JISA/CBD framework implies that from the point of view of the model, the ith signal in datasets $k = 1 : D$ is one-dimensional. By assumption (A7), we conclude that $m_i = 1$. ☐

In the rest of this section, we prove Theorem V.2. In the proof of Theorem V.2, we distinguish between two types of cases, as we now explain.

Case 1. Assume, w.l.o.g., that for some pair (i,j) we can write $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ as in (16), with the smallest D satisfying $1 \leq D < K$. Clearly, if (16) holds not only for this specific pair of (i,j) but $\forall i,j$ for the same value of D and for the same ordering of the datasets, our original JISA problem factorizes into two disjoint JISA problems that should be handled separately, one for mixtures 1 to D, and the other for $D+1$ to K. To avoid this trivial factorization, we assume that this situation does not occur. Now, (16) implies that the signal covariance matrices in datasets 1 to D, $(\mathbf{S}_{ii}^{[1:D,1:D]}, \mathbf{S}_{jj}^{[1:D,1:D]})$, are independent of $(\mathbf{S}_{ii}^{[D+1:K,D+1:K]}, \mathbf{S}_{jj}^{[D+1:K,D+1:K]})$. Therefore, the identifiability of the signal spaces of the ith and jth signals in datasets 1 to D, $\{\mathrm{span}(\mathbf{A}_i^{[k]})\}_{k=1}^{D}$ and $\{\mathrm{span}(\mathbf{A}_j^{[k]})\}_{k=1}^{D}$, depends only on the mixtures indexed by 1 to D. Therefore, w.l.o.g., we now focus on these mixtures. To simplify our discussion, we assume that $\mathbf{S}_{ii}^{[1:D,1:D]}$ and $\mathbf{S}_{jj}^{[1:D,1:D]}$ do not contain zeros. Correlations that are zero matrices will be handled soon, in Case 2. Given this assumption, we now apply Theorem V.1 to this smaller JISA problem that consists only of datasets 1 to D. By Theorem V.1, if $(\mathbf{S}_{ii}^{[1:D,1:D]}, \mathbf{S}_{jj}^{[1:D,1:D]})$ satisfy (14), $\{\mathrm{span}(\mathbf{A}_i^{[k]})\}_{k=1}^{D}$ cannot be distinguished from $\{\mathrm{span}(\mathbf{A}_j^{[k]})\}_{k=1}^{D}$. By Definitions II.1 and II.2, in this case, the overall JISA model is not identifiable.

Based on Case 1, we obtain the following result, which proves Scenario 2 in Theorem V.2.

Proposition V.4. If the direct sum structure in (16) holds with $D = 1$ for some pair (i,j), the JISA model is not identifiable.

Proof. Case 1 states that if, w.l.o.g., (16) holds with $D = 1$, we should apply Theorem V.1 to the JISA sub-problem that now consists only of the single mixture indexed by $k = 1$. Proposition V.3 states that in this case, $m_i = 1 = m_j$. Since any two non-zero scalars are always proportional, (14) always holds and, by Theorem V.1, the overall JISA model is not identifiable. ☐

The following result, together with Case 2, proves Scenario 1 in Theorem V.2, and therefore concludes our proof of Theorem V.2.

Case 2. Assume that for some pair (i,j), $j \neq i$, there exists a pair (k',l') such that $\mathbf{S}_{ii}^{[k',l']} = \mathbf{0}$ and $\mathbf{S}_{jj}^{[k',l']} \neq \mathbf{0}$. Then, these $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ cannot satisfy (14) with nonsingular $\mathbf{\Psi}^{[k]}$ $\forall k$. If, however, all zero blocks in $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ are in the same locations, they do not impose any constraints, and thus the equivalence relation (14) may still hold.

We further discuss and exemplify Theorem V.2 in Sections VI and VII.

VI. Discussion of JISA Identifiability Results – Interpretation of Theorems V.1 and V.2

We identify in Theorems V.1 and V.2 two types of scenarios associated with non-identifiability. Scenario 1 in Theorem V.2 implies that the model is not identifiable if in $D \geq 2$ datasets there exist at least two independent random processes of the same size (i.e., $m_i = m_j$) that are not correlated with any of the other $(K-D)$ datasets, and whose covariances satisfy (14) or (17). One can readily verify that the transformations in (14) and (17) represent an equivalence relation. Therefore, the type of non-identifiability in Theorem V.1 and Scenario 1 in Theorem V.2 is associated with equivalence relations between covariances. A similar notion of equivalence has already been observed by [17] in the one-dimensional case, of IVA. The use of equivalence relations in formalizing the identifiability of blind estimation problems is well known (e.g., [51]).

If we define "diversity" as any property of the data that contributes to identifiability (e.g., [13]), then, since pairs of multivariate random processes with different dimensions (i.e., $m_i \neq m_j$) are always identifiable in the JISA framework, we suggest regarding block-based models with distinct multivariate (or block) dimensions as containing more "diversity" w.r.t. models in which all random processes (or blocks) have the same dimension.

Scenario 2 in Theorem V.2 implies that the model is not identifiable if in $D \geq 1$ datasets there exist at least two independent random variables that are not correlated with any random variables in any of the other $(K-1)$ datasets. It follows from Proposition V.3 that for any pair (i,j) satisfying Scenario 2 in Theorem V.2, $m_i = 1 = m_j$. This scenario can be regarded as a multiset parallel to the well-known result in classical BSS, that one cannot separate random processes when more than one of them is i.i.d. Gaussian, without additional constraints (e.g., [7]).

From a data fusion perspective, Theorems V.1 and V.2 motivate using all the available datasets, as this may reduce the risk of non-identifiable scenarios.

VII. Illustration and Validation

The following examples provide some further theoretical insight into Theorems V.1 and V.2. Theorems V.1 and V.2 were stated with emphasis on the statistically-motivated JISA model. The following examples emphasize the role of the algebraic aspect of our results, via CBD. In all the following examples, we assume that the data satisfies assumptions (A1) to (A7), unless specified otherwise.

Example VII.1. In this example, we show that if the equivalence relation (14) holds, achieving exact CBD (9) does not necessarily result in signal separation. Let the number of independent elements in each mixture be $R = 2$ such that

$$\mathbf{S}^{[k,l]} \triangleq \begin{bmatrix} \mathbf{S}_{11}^{[k,l]} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22}^{[k,l]} \end{bmatrix} \quad \forall k,l \tag{19}$$

and assume that (14) holds. Let

$$\mathbf{B}^{[k]} \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I} & -\mathbf{\Psi}^{-[k]} \\ \mathbf{I} & \mathbf{\Psi}^{-[k]} \end{bmatrix} \mathbf{A}^{-[k]} \quad \forall k\,. \tag{20}$$

Using (8) and (20), we achieve exact CBD:

$$\mathbf{B}^{[k]}\mathbf{X}^{[k,l]}\mathbf{B}^{[l]\top} = \begin{bmatrix} \mathbf{S}_{11}^{[k,l]} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{11}^{[k,l]} \end{bmatrix} \neq \begin{bmatrix} \mathbf{S}_{11}^{[k,l]} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{22}^{[k,l]} \end{bmatrix} \quad \forall k,l \tag{21}$$

where the RHS of (21) has the same block-diagonal structure as (19), even though the bottom right block of (19) is different than that of (21). By Definition II.2, the CBD of $\{\mathbf{X}^{[k,l]}\}$ is not unique. In this case,

$$\mathbf{B}^{[k]}\mathbf{x}^{[k]} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{s}_1^{[k]} - \mathbf{\Psi}^{-[k]}\mathbf{s}_2^{[k]} \\ \mathbf{s}_1^{[k]} + \mathbf{\Psi}^{-[k]}\mathbf{s}_2^{[k]} \end{bmatrix} \quad \forall k \tag{22}$$

which does not separate $\mathbf{s}_1^{[k]}$ from $\mathbf{s}_2^{[k]}$ $\forall k$, and hence, this JISA model is not identifiable. One can readily verify that this example equally hold if we replace m_i with $m_i^{[k]}$ $\forall i,k$. We mention that a structure similar to (20) of an unmixing matrix was used in [52, Chapter 4.5], in an example for non-identifiability of piecewise stationary ISA. In Section VIII, we elaborate on the link between these two models.

Example VII.2. In this example, we illustrate Scenario 2 in Theorem V.2. Let $K = 3$, $R \geq 3$, and

$$\mathbf{S}_{ii} = \begin{bmatrix} \mathbf{S}_{ii}^{[1,1]} & \mathbf{0} & \mathbf{S}_{ii}^{[1,3]} \\ \mathbf{0} & \mathbf{S}_{ii}^{[2,2]} & \mathbf{0} \\ \mathbf{S}_{ii}^{[3,1]} & \mathbf{0} & \mathbf{S}_{ii}^{[3,3]} \end{bmatrix}, \; \mathbf{S}_{jj} = \begin{bmatrix} \mathbf{S}_{jj}^{[1,1]} & \mathbf{0} & \mathbf{S}_{jj}^{[1,3]} \\ \mathbf{0} & \mathbf{S}_{jj}^{[2,2]} & \mathbf{0} \\ \mathbf{S}_{jj}^{[3,1]} & \mathbf{0} & \mathbf{S}_{jj}^{[3,3]} \end{bmatrix} \tag{23}$$

for some $i \neq j$. This setup is not identifiable, because we can permute the block rows and columns of $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ using the same permutation to obtain the structure in (18) with $D = 1$ (by Proposition V.3, we also have $m_i = m_j = 1$). However, if we replace one of these zero blocks with a nonsingular matrix (recall symmetry), we obtain a setup that is identifiable by Theorem V.2.

Example VII.3. The following structure, with $K = 3$ and $R = 2$, does not satisfy any scenario in Theorem V.2 and is thus always identifiable.

$$\mathbf{S}_{ii} = \begin{bmatrix} \mathbf{S}_{ii}^{[1,1]} & \mathbf{S}_{ii}^{[1,2]} & \mathbf{0} \\ \mathbf{S}_{ii}^{[2,1]} & \mathbf{S}_{ii}^{[2,2]} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{ii}^{[3,3]} \end{bmatrix}, \; \mathbf{S}_{jj} = \begin{bmatrix} \mathbf{S}_{jj}^{[1,1]} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_{jj}^{[2,2]} & \mathbf{S}_{jj}^{[2,3]} \\ \mathbf{0} & \mathbf{S}_{jj}^{[3,2]} & \mathbf{S}_{jj}^{[3,3]} \end{bmatrix} \tag{24}$$

Furthermore, by Proposition V.3, the blocks of these matrices are scalars, i.e., $m_i = 1 = m_j$.

VIII. Discussion

In this section, we discuss several implications of Theorem V.1, in a broader context.

A. A Link between JISA and ISA

There exist various types of links between JISA and ISA. In this section, we focus only on one of these links. More specifically, we now show that previous results on the identifiability of piecewise stationary ISA can be obtained as a special case of Theorem V.1. In what follows, we assume that the data satisfies assumptions (A1) to (A7).

Let all $\mathbf{A}^{[k]}$ be identical, i.e., $\mathbf{A}^{[k]} \triangleq \mathbf{A} \in \mathbb{R}^{M\times M}$ $\forall k$. Let each pair of indices (k,l) be mapped into a single index, i.e., $(k,l) \mapsto q$. In this case, (8) and (9) rewrite, respectively, as

$$\mathbf{X}^{(q)} = \mathbf{A}\mathbf{S}^{(q)}\mathbf{A}^\top = \sum_{i=1}^{R} \mathbf{A}_i \mathbf{S}_{ii}^{(q)} \mathbf{A}_i^\top \quad q = 1,\dots,Q \tag{25}$$

and

$$\mathbf{A}^{-1}\mathbf{X}^{(q)}\mathbf{A}^{-\top} = \mathbf{S}^{(q)} \quad q = 1,\dots,Q \tag{26}$$

where $\mathbf{S}^{(q)} \triangleq \oplus_{i=1}^{R} \mathbf{S}_{ii}^{(q)}$, $\mathbf{S}_{ii}^{(q)} \in \mathbb{R}^{m_i\times m_i}$, $\mathbf{A} = [\mathbf{A}_1|\cdots|\mathbf{A}_R]$, $\mathbf{A}_i \in \mathbb{R}^{M\times m_i}$, and Q is the number of distinct equations in (26). The factorization in (26) is often referred to as JBD (e.g., [25]). In analogy to (10), each summand in (25) remains invariant if a pair $(\mathbf{A}_i, \mathbf{S}_{ii}^{(q)})$ is replaced with $(\mathbf{A}_i\mathbf{Z}_{ii}^{-1}, \mathbf{Z}_{ii}\mathbf{S}_{ii}^{(q)}\mathbf{Z}_{ii}^\top)$ for an arbitrary nonsingular $\mathbf{Z}_{ii}$. In analogy to Definition II.2 (see also, e.g., [19], [46]),

Definition VIII.1. If, for given $\{\mathbf{X}^{(q)}\}_{q=1}^{Q}$, and given $\{m_i\}_{i=1}^{R}$, any choice of $\mathbf{A}$ and $\{\mathbf{S}^{(q)}\}_{q=1}^{Q}$ that satisfy (25) yields the same R summands $\mathbf{A}_i\mathbf{S}_{ii}^{(q)}\mathbf{A}_i^\top$ $\forall q$, we say that the decomposition in (25) (and the JBD (26), if applicable) is essentially unique.

Alternatively, we can restate Definition VIII.1 by replacing the R summands with the R m_i-dimensional signal subspaces $\mathrm{span}(\mathbf{A}_i)$. Similarly, Definition III.1 simplifies as:

Definition VIII.2. A sequence $\{\mathbf{S}_{ii}^{(q)}\}_{q=1}^{Q}$ of $m_i\times m_i$ matrices is said to be *reducible by simultaneous congruence* if there exists a transformation (nonsingular matrix) $\mathbf{T} \in \mathbb{R}^{m_i\times m_i}$ such that

$$\mathbf{T}\mathbf{S}_{ii}^{(q)}\mathbf{T}^\top = \begin{bmatrix} \mathbf{S}_{i_1i_1}^{(q)} & \mathbf{0} \\ \mathbf{0}_{\alpha\times\beta} & \mathbf{S}_{i_2i_2}^{(q)} \end{bmatrix} \quad \forall q\,, \tag{27}$$

where α and $\beta \triangleq m_i - \alpha$ are positive integers. Otherwise, the sequence is said to be *irreducible by simultaneous congruence*.

One can readily verify that if $\mathbf{S}_{ii}$ is irreducible by coupled congruence (Definition III.1), then the sequence $\mathbf{S}_{ii}^{[1,1]}, \mathbf{S}_{ii}^{[1,2]}, \dots, \mathbf{S}_{ii}^{[K,K]}$ consisting of its blocks is irreducible by simultaneous congruence (Definition VIII.2). However, the converse does not necessarily hold. Applying the same simplification procedure to (14), such that $\mathbf{\Psi}^{[k]} \triangleq \mathbf{\Phi} \in \mathbb{R}^{m_i\times m_i}$ is an arbitrary nonsingular matrix $\forall k$, we obtain

$$\mathbf{S}_{jj}^{(q)} = \mathbf{\Phi}\mathbf{S}_{ii}^{(q)}\mathbf{\Phi}^\top \quad \forall q\,. \tag{28}$$

Equation (28) is an equivalence relation between $\{\mathbf{S}_{ii}^{(q)}\}_{q=1}^{Q}$ and $\{\mathbf{S}_{jj}^{(q)}\}_{q=1}^{Q}$. Next, due to the symmetry of the congruence transformation, we impose symmetry on $\mathbf{X}^{(q)}$ and $\mathbf{S}^{(q)}$ $\forall q$. Finally, in accordance with Theorem V.1, the nonsingularity of $\mathbf{S}_{ii}^{[k,l]}$ translates into assuming that $\mathbf{S}^{(q)}$, and thus also $\mathbf{X}^{(q)}$, are positive-definite. Given these assumptions, one can readily verify that (26) summarizes the sufficient statistics of ISA, when each of the underlying signals is a piecewise stationary multivariate (m_i-variate) Gaussian process with uncorrelated samples. In this case, $\mathbf{S}_{ii}^{(q)}$ is the covariance of the ith random vector in the qth stationary interval [26], [47].

Applying all these simplifications to Theorem V.1, we obtain the following theorem:

Theorem VIII.3. *Consider an ISA model whose sufficient statistics are given by (25), where $\mathbf{A} \in \mathbb{R}^{M\times M}$ is a nonsingular matrix, and $\{\mathbf{S}_{ii}^{(q)}\}_{q=1}^{Q}$ a sequence of positive-definite real-valued symmetric $m_i\times m_i$ matrices, irreducible by simultaneous congruence (Definition VIII.2), for any $i = 1,\dots,R$. Then this ISA is not identifiable, and the JBD (26) is not unique iff there exists at least one pair (i,j) that satisfies (28) for some nonsingular $m_i\times m_i$ matrix $\mathbf{\Phi}$.*

The key point is that Theorem VIII.3, which we have just obtained by simplifying Theorem V.1, is indeed identical to the theorem on the identifiability of ISA, and the uniqueness of JBD, as previously derived in [19] (see also [26], [27]). To conclude, we have shown that the uniqueness and identifiability of JBD (26) and of piecewise stationary ISA can be regarded as special cases of the uniqueness and identifiability of CBD (9) and stationary JISA.

B. Implications on Block-Based Decompositions

The Q matrices $\{\mathbf{X}^{(q)}\}$ can be stored in a single three-dimensional $M\times M\times Q$ array, sometimes referred to as a third-order tensor. In this case, the decomposition in (25) is nothing but a special case of a decomposition of a tensor in a sum of R low-rank block terms [46]. BTD is a class of tensor decompositions that attract increasing interest, due to their ability to model various latent structures in data. Accordingly, Theorem VIII.3 is a special case of more general results on the uniqueness of BTD. The decompositions in (8b) and (9) cannot be written compactly as a decomposition of single tensor in a sum of low-rank terms. This is not a flaw: this is simply because our model is more general. The fact that certain results on the uniqueness of tensor decompositions are special cases of results on the uniqueness of a more general class of coupled block decompositions (8) hints that other fundamental concepts in the analysis of the uniqueness of tensor decompositions, such as Kruskal's rank [53], generalize as well. We point out that the concept of Kruskal's rank was generalized in [46] for matrices partitioned column-wise into blocks, however, not for block elements such as $\{\mathbf{S}_{ii}^{(q)}\}_{q=1}^{Q}$ in (25), which are $m_i\times m_i\times Q$ tensors and not matrices. We leave further discussion of this topic to future publications.

C. Can JISA Identifiability be Obtained from the GEVD?

In Section IV, we mentioned that when $K = 2$, we can write JISA as GEVD, and if the generalized eigenvalues are distinct, then the one-dimensional subspaces associated with each column of $\mathbf{A}^{[1]}$ and $\mathbf{A}^{[2]}$ are guaranteed to be unique. One may wonder if we can use this result to obtain the uniqueness results of JISA with $K \geq 3$ datasets, by observing the generalized eigenvalues, for each pair of datasets. The answer is as follows. Indeed, if all the pairwise generalized eigenvalues are distinct, then all the one-dimensional subspaces are distinct and unique, and therefore, also if we collect them into subspaces of a higher dimension, these higher-dimensional subspaces will remain distinct and thus uniquely identifiable. Therefore, this is a sufficient condition. However, when these generalized eigenvalues are not distinct, we can still have identifiability: this follows directly from Equation (14). One can readily verify that even if each pair of datasets has some non-distinct generalized eigenvalues, (14) will not necessarily be satisfied, as soon as $K \geq 3$, and $m_i \geq 2$ for at least one i.

D. Noise

We now briefly discuss our identifiability results in the presence of additive noise, when assumptions (A1) to (A7) hold. A prerequisite for identifiability is that the constraints imposed by the observations be no fewer than the number of free model parameters. In the noise free case, the inequality was strict at $K = 2$, and satisfied for $K \geq 2$, for any value of m_i [1]. Clearly, when the noise parameters have to be estimated as well, the case $K = 2$ is never identifiable. Furthermore, the results now depend both on the specific noise model, as well as on the values of m_i. Hence, in the noisy case, this prerequisite has to be tested individually for each noisy JISA model. As we already know from our analysis in the previous sections, a sufficient number of d.o.f. does not guarantee identifiability, and, as we have shown in this paper, characterizing all the cases of non-identifiability is not a simple task, even in the absence of noise. Nevertheless, it is easy to see that additive noise cannot improve the identifiability of the model: in the presence of additive noise, $\mathbf{x}^{[k]} = \mathbf{A}^{[k]}\mathbf{s}^{[k]} + \mathbf{n}^{[k]}$ $\forall k$, $\mathbf{n}^{[k]}$ being the noise random process, (22) will have an additive term $\mathbf{B}^{[k]}\mathbf{n}^{[k]}$ on the RHS $\forall k$. Clearly, the non-identifiable mixed sources with arbitrary $\mathbf{\Psi}^{[k]}$ remain non-identifiable and mixed. The same conclusion will apply even if we use another filtering approach with better signal to noise ratio (SNR), instead of $\mathbf{B}^{[k]}$. It remains to see whether the noise can add non-identifiable cases when the noise-free model is identifiable by Theorems V.1 and V.2. This question cannot be answered using our existing results, and requires a new analysis, for example, by derivation of the FIM explicitly for the noisy case (see, e.g., [54], [55]). Such derivations are beyond the scope of this paper. Algorithms and analysis for JISA, in the presence of perturbation due to finite sample size and/or additive noise, are described, e.g., in [1], [2], [10], [11].

IX. Conclusion

In this paper, we fully characterized the uniqueness and identifiability of JISA in a setup in which each dataset is not identifiable individually. We proved that this JISA model is generally identifiable, except when the SOS of two or more of its underlying sources belong to the same equivalence class. Since two multivariate random processes that have different dimensions cannot satisfy this equivalence relation, their presence implies that they are always identifiable within our JISA framework. This result implies that the presence of terms of different dimensions enhances identifiability, and thus can be regarded as a new type of diversity in the data. This result further motivates models in which the data can be represented in block form, instead of rank-1 elements. We explained how this result generalizes and extends known results on the identifiability and uniqueness of nonstationary ISA and JBD [19]. We conjectured that insights from these results can be applied to more general types of block-based decompositions, and, in particular, extend the concept of Kruskal's rank to more elaborate types of coupled matrix and tensor block decompositions. From a data fusion perspective, we provided new theoretical evidence that a link among datasets can achieve uniqueness and identifiability in cases where each dataset is not unique or identifiable individually. Our results provide further evidence that the concept of irreducibility is a key factor in subspace-based BSS and in decompositions in sum of low-rank block terms. We have shown that analysing the identifiability of new signal processing models that are inspired by data fusion leads to the development of new theoretical results. These include the identities in Appendix A on block partitioned matrices. The solutions to the system of coupled matrix equations in (59) were derived in a separate publication [23], [24].

Appendix A
Some Algebraic Properties

For any matrices $\mathbf{A}, \mathbf{B}, \mathbf{X}, \mathbf{Y}$ (with appropriate dimensions),

$$(\mathbf{B} \otimes \mathbf{A})(\mathbf{Y} \otimes \mathbf{X}) = \mathbf{BY} \otimes \mathbf{AX} \tag{29a}$$
$$(\mathbf{B} \otimes \mathbf{A})^\top = \mathbf{B}^\top \otimes \mathbf{A}^\top \tag{29b}$$
$$\mathrm{vec}(\mathbf{AXB}^\top) = (\mathbf{B} \otimes \mathbf{A})\mathrm{vec}(\mathbf{X}) \tag{29c}$$

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

$$\mathfrak{T}_{M,N}(\mathbf{B} \otimes \mathbf{A}) = (\mathbf{A} \otimes \mathbf{B})\mathfrak{T}_{P,Q} \tag{30}$$

where the *commutation matrix* $\mathfrak{T}_{P,Q} \in \mathbb{R}^{PQ\times PQ}$ satisfies

$$\mathrm{vec}(\mathbf{M}^\top) = \mathfrak{T}_{P,Q}\mathrm{vec}(\mathbf{M}) \tag{31}$$

for any $P \times Q$ matrix $\mathbf{M}$ [56]. Equations (29) and (30) can be found, e.g., in [56], [57].

Definition A.1 (vecd Operator)**.** For any square matrix $\mathbf{X}$ of size $K \times K$ with entries $x_{kk'}$, where $k, k' = 1, \ldots, K$,

$$\mathrm{vecd}(\mathbf{X}) \triangleq \begin{bmatrix} x_{11} & \cdots & x_{KK} \end{bmatrix}^\top . \tag{32}$$

The $\mathrm{vecd}(\cdot)$ operator can be found, e.g., in [58, Eq. (7)].

Definition A.2 (vecbd Operator)**.** For any matrix $\mathbf{X}$ of size $\alpha \times \beta$, partitioned into $K \times K$ blocks such that its (k, k')th block $\mathbf{X}_{kk'}$ has size $\alpha_k \times \beta_{k'}$, where $k, k' = 1, \ldots, K$, $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]^\top$, $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_K]^\top$, $\alpha = \sum_{k=1}^K \alpha_k$, and $\beta = \sum_{k=1}^K \beta_k$,

$$\mathrm{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\} \triangleq \begin{bmatrix} \mathrm{vec}^\top(\mathbf{X}_{11}) & \cdots & \mathrm{vec}^\top(\mathbf{X}_{KK}) \end{bmatrix}^\top . \tag{33}$$

$\mathrm{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\}$ is a vector of length $\boldsymbol{\alpha}^\top\boldsymbol{\beta}$, consisting 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 $\mathrm{vecbd}_{\boldsymbol{\alpha}}\{\mathbf{X}\} \triangleq \mathrm{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\alpha}}\{\mathbf{X}\}$. The $\mathrm{vecbd}_{\boldsymbol{\alpha}}\{\cdot\}$ operator can be found, e.g., in [59].

Identity A.3.

$$(\mathbf{A} \boxbar \mathbf{B})^\top(\mathbf{C} \boxbar \mathbf{D}) = \mathbf{A}^\top\mathbf{C} \boxplus \mathbf{B}^\top\mathbf{D}\,. \tag{34}$$

for any $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$, such that all products are defined.

Proof. Let $\mathbf{A}$ and $\mathbf{B}$ be partitioned column-wise in K blocks, and $\mathbf{C}$ and $\mathbf{D}$ in L blocks. Using (29b) we can write

$$(\mathbf{A} \boxbar \mathbf{B})^\top(\mathbf{C} \boxbar \mathbf{D}) = \left[\begin{array}{c} \mathbf{A}_1^\top \otimes \mathbf{B}_1^\top \\ \hline \vdots \\ \hline \mathbf{A}_K^\top \otimes \mathbf{B}_K^\top \end{array}\right] \left[\begin{array}{c|c|c} \mathbf{C}_1 \otimes \mathbf{D}_1 & \cdots & \mathbf{C}_L \otimes \mathbf{D}_L \end{array}\right] \tag{35}$$

whose (k,l)th block $\mathbf{A}_k^\top\mathbf{C}_l \otimes \mathbf{B}_k^\top\mathbf{D}_l$, obtained via (29a), is exactly the (k,l)th block of $\mathbf{A}^\top\mathbf{C} \boxplus \mathbf{B}^\top\mathbf{D}$. □

Remark A.4. In [18, Identity 6.1], we introduced a special case of Identity A.3:

$$(\mathbf{A} \odot \mathbf{B})^\top(\mathbf{C} \odot \mathbf{D}) = \mathbf{A}^\top\mathbf{C} * \mathbf{B}^\top\mathbf{D} \tag{36}$$

The more familiar identity (e.g., [60]–[63]),

$$\mathbf{A}^\top\mathbf{A} * \mathbf{B}^\top\mathbf{B} = (\mathbf{A} \odot \mathbf{B})^\top(\mathbf{A} \odot \mathbf{B})\,, \tag{37}$$

is thus a special case of Identity A.3 and [18, Identity 6.1].

Identity A.5. Let $\mathbf{A} \in \mathbb{R}^{\mu\times\alpha}$ and $\mathbf{B} \in \mathbb{R}^{\nu\times\beta}$ be two matrices partitioned column-wise into K blocks of dimensions $\mu\times\alpha_k$ and $\nu\times\beta_k$, respectively, $\alpha = \sum_{k=1}^K \alpha_k$, $\beta = \sum_{k=1}^K \beta_k$, $\boldsymbol{\alpha} = [\alpha_1,\ldots,\alpha_K]^\top$, $\boldsymbol{\beta} = [\beta_1,\ldots,\beta_K]^\top$, as follows,

$$\begin{aligned} \mathbf{A} &= \left[\begin{array}{c|c|c} \mathbf{A}_1 & \cdots & \mathbf{A}_K \end{array}\right], \quad \mathbf{A}_k \in \mathbb{R}^{\mu\times\alpha_k} \\ \mathbf{B} &= \left[\begin{array}{c|c|c} \mathbf{B}_1 & \cdots & \mathbf{B}_K \end{array}\right], \quad \mathbf{B}_k \in \mathbb{R}^{\nu\times\beta_k} \end{aligned} \tag{38}$$

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

$$(\mathbf{B} \boxbar \mathbf{A})\mathrm{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\mathbf{X}\} = \mathrm{vec}(\mathbf{A}\mathbf{X}\mathbf{B}^\top)\,, \tag{39}$$

where the operator "$\mathrm{vecbd}_{\boldsymbol{\alpha}\times\boldsymbol{\beta}}\{\cdot\}$" was defined in Definition A.2.

Proof. Identity A.5 is a special case of [59, Theorem 4.17], when the partition is only column-wise. Alternatively, set $K=1$ and $N=1$ in [46, Equation (2.14)]. □

If we set $K=1$ in Identity A.5, that is, $\mathbf{X}$ has only one block, we obtain (29c). If we set $\alpha_k = 1 = \beta_k\ \forall k$ in Identity A.5, that is, $\mathbf{X}$ is a diagonal matrix, we obtain

$$(\mathbf{B} \odot \mathbf{A})\mathrm{vecd}(\mathbf{X}) = (\mathbf{B} \otimes \mathbf{A})\mathrm{vec}(\mathbf{X}) = \mathrm{vec}(\mathbf{A}\mathbf{X}\mathbf{B}^\top) \tag{40}$$

where the RHS of (40) is by (29c) and the operator "$\mathrm{vecd}(\cdot)$" was defined in Definition A.1. Identity (40) can be found, e.g., in [58, Table III, T3.13] and [60, Equation (27)].

We briefly explain the link to tensor decomposition. Our proof of Identity A.5 states that (39) is a vector representation of a decomposition in terms with multilinear rank (α_r,β_r) [46] of a second-order tensor (i.e., a matrix). Identities (29c) and (40) are thus two special cases thereof: (29c) and (40) can be regarded, respectively, as a vectorization of a decomposition in sum of rank-1 terms, and as a vectorization of a Tucker format [64], of a second-order tensor. This is not surprising, because both types of tensor factorizations are special cases of the decomposition in low multilinear rank terms [46].

Appendix B
Analyzing the Singularity of the FIM

In [1], we have shown (see Footnote 3) that asymptotically, that is, when the number of samples drawn from the random variables goes to infinity, for every pair (i,j) with $i \neq j$, the estimation error of the parameters in the model that we have just defined is proportional to the inverse of the symmetric positive semi-definite $2\mathbf{m}_i^\top\mathbf{m}_j \times 2\mathbf{m}_i^\top\mathbf{m}_j$ matrix

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1} & \oplus_{k=1}^K \boldsymbol{\mathcal{T}}_{m_j^{[k]},m_i^{[k]}} \\ \oplus_{k=1}^K \boldsymbol{\mathcal{T}}_{m_i^{[k]},m_j^{[k]}} & \mathbf{S}_{ii} \boxplus \mathbf{S}_{jj}^{-1} \end{bmatrix} \tag{41}$$

where the commutation matrix $\boldsymbol{\mathcal{T}}_{P,Q} \in \mathbb{R}^{PQ\times PQ}$ was defined in (31), and $\mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1}$ is an $\mathbf{m}_i^\top\mathbf{m}_j \times \mathbf{m}_i^\top\mathbf{m}_j$ matrix whose (k,l)th block has size $m_i^{[k]}m_j^{[k]} \times m_i^{[l]}m_j^{[l]}$. Using Identity (30) in Appendix A, we can write

$$\boldsymbol{\mathcal{H}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \oplus_{k=1}^K \boldsymbol{\mathcal{T}}_{m_i^{[k]},m_j^{[k]}} \end{bmatrix} \mathbf{H} \begin{bmatrix} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \oplus_{k=1}^K \boldsymbol{\mathcal{T}}^\top_{m_i^{[k]},m_j^{[k]}} \end{bmatrix} \tag{42}$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{S}_{jj} \boxplus \mathbf{S}_{ii}^{-1} & \mathbf{I} \\ \mathbf{I} & \mathbf{S}_{jj}^{-1} \boxplus \mathbf{S}_{ii} \end{bmatrix}. \tag{43}$$

Matrix $\mathbf{H}$ is always well-defined because it is derived (in [1]) based on the assumption that $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ are positive-definite covariance matrices. Matrix $\boldsymbol{\mathcal{H}}$ and its inverse are the main ingredients in the closed-form expression for the CRLB and FIM, as explained in [1]. Therefore, model identifiability boils down to characterizing all the cases in which $\mathbf{H}$ is singular.

For $\mathbf{H}$ to be positive-definite[4], we require that for any vector $\mathbf{x}$ of length $2\mathbf{m}_i^\top\mathbf{m}_j$,

$$0 < \mathbf{x}^\top\mathbf{H}\mathbf{x} = \mathbf{x}^\top\mathbf{V}^\top\mathbf{V}\mathbf{x} \tag{44}$$

where $\mathbf{V}$ is such that $\mathbf{H} = \mathbf{V}^\top\mathbf{V}$. Conversely, for $\mathbf{H}$ not to be positive-definite, there must exist some non-zero vector $\mathbf{x}$ of length $2\mathbf{m}_i^\top\mathbf{m}_j$ such that

$$0 = \mathbf{x}^\top\mathbf{H}\mathbf{x} = \mathbf{x}^\top\mathbf{V}^\top\mathbf{V}\mathbf{x} \tag{45a}$$

$$\Leftrightarrow\ \mathbf{V}\mathbf{x} = \mathbf{0}\,. \tag{45b}$$

Based on (44) and (45), we now look for a meaningful factorization of the Gram matrix $\mathbf{H} = \mathbf{V}^\top\mathbf{V}$. We propose

$$\mathbf{H} = \begin{bmatrix} (\mathbf{S}_{jj}^{\frac{1}{2}\top} \boxbar \mathbf{S}_{ii}^{-\frac{1}{2}})^\top \\ (\mathbf{S}_{jj}^{-\frac{1}{2}} \boxbar \mathbf{S}_{ii}^{\frac{1}{2}\top})^\top \end{bmatrix} \begin{bmatrix} \mathbf{S}_{jj}^{\frac{1}{2}\top} \boxbar \mathbf{S}_{ii}^{-\frac{1}{2}} & \mathbf{S}_{jj}^{-\frac{1}{2}} \boxbar \mathbf{S}_{ii}^{\frac{1}{2}\top} \end{bmatrix} \tag{46}$$

where

$$\mathbf{V} \triangleq \begin{bmatrix} \mathbf{S}_{jj}^{\frac{1}{2}\top} \boxbar \mathbf{S}_{ii}^{-\frac{1}{2}} & \mathbf{S}_{jj}^{-\frac{1}{2}} \boxbar \mathbf{S}_{ii}^{\frac{1}{2}\top} \end{bmatrix} \tag{47}$$

[4]The derivations in Appendix B are valid also for complex-valued variables, when $^\top$ is replaced with $^\mathsf{H}$, wherever applicable.

has size $(\mathbf{m}_i^\top \mathbf{1}_K)(\mathbf{m}_j^\top \mathbf{1}_K) \times 2\mathbf{m}_i^\top \mathbf{m}_j$. Matrix $\mathbf{S}_{ii}^{\frac{1}{2}}$ is defined such that $\mathbf{S}_{ii} = \mathbf{S}_{ii}^{\frac{1}{2}} \mathbf{S}_{ii}^{\frac{1}{2}\top}$; this factorization may be obtained, e.g., via SVD. The equality in (46) follows from Identity A.3 in Appendix A. A prerequisite for the nonsingularity of $\mathbf{H}$ is that the number of rows of $\mathbf{V}$ be equal to or larger than the number of its columns. One can readily verify that when assumption (A7) holds, this condition is satisfied as soon as $K \geq 2$, because $(\mathbf{m}_i^\top \mathbf{1}_K)(\mathbf{m}_j^\top \mathbf{1}_K) = Km_i \cdot Km_j \geq 2Km_i m_k = 2\mathbf{m}_i^\top \mathbf{m}_j$. The fact that this does not hold for $K = 1$ is yet another proof that a single dataset is not identifiable. Otherwise, satisfying this inequality depends on the specific sizes of the blocks. For example, if $\mathbf{m}_i = [1, 1, 5]^\top = \mathbf{m}_j$, this inequality does not hold, and this case is not identifiable. In what follows, we consider only data whose dimensions satisfy the desired inequality. In these cases, the rank of $\mathbf{H}$ may be equal to $2\mathbf{m}_i^\top \mathbf{m}_j$. In this paper, our goal is thus to determine the additional structural conditions on $\mathbf{H}$ and $\mathbf{V}$ such that this rank will be smaller, in the special case where assumptions (A1) to (A7) hold.

Next, we look for a non-zero vector $\mathbf{x}$ such that $\mathbf{V}\mathbf{x} = \mathbf{0}$. W.l.o.g., we look for $\mathbf{x}$ in the form

$$\mathbf{x} = \begin{bmatrix} \boldsymbol{\mu}^\top & -\boldsymbol{\nu}^\top \end{bmatrix}^\top . \tag{48}$$

Substituting this in (45b), we obtain

$$\begin{bmatrix} \mathbf{S}_{jj}^{\frac{1}{2}\top} \boxbar \mathbf{S}_{ii}^{-\frac{1}{2}} & \mathbf{S}_{jj}^{-\frac{1}{2}} \boxbar \mathbf{S}_{ii}^{\frac{1}{2}\top} \end{bmatrix} \begin{bmatrix} \boldsymbol{\mu} \\ -\boldsymbol{\nu} \end{bmatrix} = \mathbf{0} \tag{49}$$

for some non-zero $\boldsymbol{\mu}$ and/or $\boldsymbol{\nu}$. We now turn to finding these $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$. As we shall see soon, it is useful to rearrange the elements of $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ in the following structure:

$$\boldsymbol{\mu} \triangleq \begin{bmatrix} \mathrm{vec}^\top(\mathbf{M}^{[1]}) & \cdots & \mathrm{vec}^\top(\mathbf{M}^{[K]}) \end{bmatrix} \tag{50a}$$
$$\boldsymbol{\nu} \triangleq \begin{bmatrix} \mathrm{vec}^\top(\mathbf{N}^{[1]}) & \cdots & \mathrm{vec}^\top(\mathbf{N}^{[K]}) \end{bmatrix} \tag{50b}$$

where the size of each $\mathbf{M}^{[k]}$ and $\mathbf{N}^{[k]}$ is $m_i^{[k]} \times m_j^{[k]}$. Equation (50a) can be rewritten more compactly as

$$\boldsymbol{\mu} = \mathrm{vecbd}_{\mathbf{m}_i \times \mathbf{m}_j}\{\mathbf{M}\} \;, \; \boldsymbol{\nu} = \mathrm{vecbd}_{\mathbf{m}_i \times \mathbf{m}_j}\{\mathbf{N}\} \tag{51}$$

where the "vecbd" operator was described in Definition A.2 in Appendix A, $\mathbf{M} \triangleq \oplus_{k=1}^K \mathbf{M}^{[k]}$, $\mathbf{N} \triangleq \oplus_{k=1}^K \mathbf{N}^{[k]}$, and the row $\times$ column partition of both $\mathbf{M}$ and $\mathbf{N}$ is $\mathbf{m}_i \times \mathbf{m}_j$. Let us rewrite (49) as

$$(\mathbf{S}_{jj}^{\frac{1}{2}\top} \boxbar \mathbf{S}_{ii}^{-\frac{1}{2}})\boldsymbol{\mu} = (\mathbf{S}_{jj}^{-\frac{1}{2}} \boxbar \mathbf{S}_{ii}^{\frac{1}{2}\top})\boldsymbol{\nu} \,. \tag{52}$$

Applying Identity A.5 in Appendix A to (52) yields

$$\mathrm{vec}(\mathbf{S}_{ii}^{\frac{1}{2}\top} \mathbf{N} \mathbf{S}_{jj}^{-\frac{1}{2}\top}) = \mathrm{vec}(\mathbf{S}_{ii}^{-\frac{1}{2}} \mathbf{M} \mathbf{S}_{jj}^{\frac{1}{2}}) \,. \tag{53}$$

Removing the "vec" notation, (53) rewrites as

$$\mathbf{S}_{ii}^{\frac{1}{2}\top} \mathbf{N} \mathbf{S}_{jj}^{-\frac{1}{2}\top} = \mathbf{S}_{ii}^{-\frac{1}{2}} \mathbf{M} \mathbf{S}_{jj}^{\frac{1}{2}} \,. \tag{54}$$

Since $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ are positive-definite, (54) rewrites as

$$\mathbf{S}_{ii}\mathbf{N} = \mathbf{M}\mathbf{S}_{jj} \;\Leftrightarrow\; \mathbf{S}_{ii}^{[k,l]}\mathbf{N}^{[l]} = \mathbf{M}^{[k]}\mathbf{S}_{jj}^{[k,l]} \quad \forall k,l \,. \tag{55}$$

Hence, our task of finding $\boldsymbol{\mu}$ and $\boldsymbol{\nu}$ has been recast into finding $\{\mathbf{M}^{[k]}, \mathbf{N}^{[k]}\}_{k=1}^K$, not all zero, for which equality (55) holds. The matrix equation (55) can further be simplified if we normalize $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$ such that their blocks on the main diagonal are equal to the identity,

$$\widetilde{\mathbf{S}}_{ii}^{[k,k]} = \mathbf{I}_{m_i^{[k]}} \quad \text{and} \quad \widetilde{\mathbf{S}}_{jj}^{[k,k]} = \mathbf{I}_{m_j^{[k]}} \quad \forall k \,. \tag{56}$$

We do so using the transformation

$$\widetilde{\mathbf{S}}_{ii}^{[k,l]} = (\mathbf{S}_{ii}^{[k,k]})^{-\frac{1}{2}} \mathbf{S}_{ii}^{[k,l]} (\mathbf{S}_{ii}^{[l,l]})^{-\frac{1}{2}\top} \quad \forall k,l \,, \tag{57}$$

which is absorbed by the arbitrary factors discussed in Section II-A, and thus does not alter the identifiability of the JISA model in question. Substituting (57) in (55), we obtain

$$\underbrace{(\mathbf{S}_{ii}^{[k,k]})^{-\frac{1}{2}} \mathbf{M}^{[k]} (\mathbf{S}_{jj}^{[k,k]})^{\frac{1}{2}}}_{\mathbf{L}^{[k]}} \widetilde{\mathbf{S}}_{jj}^{[k,l]} = \widetilde{\mathbf{S}}_{ii}^{[k,l]} \overbrace{(\mathbf{S}_{ii}^{[l,l]\top})^{\frac{1}{2}} \mathbf{N}^{[l]} (\mathbf{S}_{jj}^{[l,l]\top})^{-\frac{1}{2}}}^{\mathbf{L}'^{[l]}} \,. \tag{58}$$

Substituting (56) in (58) for $l = k$ implies $\mathbf{L}^{[k]} = \mathbf{L}'^{[k]}$ $\forall k$. Problem (55) can now be restated as characterizing the non-trivial solutions of

$$\widetilde{\mathbf{S}}_{ii}\mathbf{L} = \mathbf{L}\widetilde{\mathbf{S}}_{jj} \;\Leftrightarrow\; \widetilde{\mathbf{S}}_{ii}^{[k,l]}\mathbf{L}^{[l]} = \mathbf{L}^{[k]}\widetilde{\mathbf{S}}_{jj}^{[k,l]} \quad \forall k,l \tag{59}$$

where $\mathbf{L}^{[k]}$ are $m_i^{[k]} \times m_j^{[k]}$ matrices and $\mathbf{L} \triangleq \oplus_{k=1}^K \mathbf{L}^{[k]}$. For fixed (k,l), the RHS of (59) is a Sylvester-type homogeneous matrix equation. When all the indices k,l are considered at once, the RHS of (59) is a system of coupled Sylvester-type homogeneous matrix equations, in K unknowns $\{\mathbf{L}^{[k]}\}_{k=1}^K$. The left-hand side (LHS) of (59) can be described as a single structured Sylvester-type homogeneous matrix equation, in a single structured unknown $\mathbf{L}$. Problem (59) is simpler than (55) because $\mathbf{L}$ replaces both $\mathbf{M}$ and $\mathbf{N}$, reducing by half the number of unknowns. The JISA identifiability problem can now be restated as characterizing the minimal set of additional conditions on $\widetilde{\mathbf{S}}_{ii}$ and $\widetilde{\mathbf{S}}_{jj}$ such that $\{\mathbf{L}^{[k]}\}_{k=1}^K$ are not all zero and (59) holds, given the rank and irreducibility constraints imposed by assumptions (A1) to (A7).

Appendix C
Non-trivial Solutions to (59)

Here, we characterize the non-trivial solutions of (59) when $\mathbf{S}_{ii}^{[k,l]} \in \mathbb{R}^{m_i \times m_i}$ are nonsingular matrices $\forall i,k,l$, and assumptions (A1) to (A7) hold. The generalization to zero-valued blocks is explained in detail in Section V. Hence, we can assume that $\mathbf{S}_{ii}$ and $\mathbf{S}_{jj}$, as well as their normalized forms $\widetilde{\mathbf{S}}_{ii}$ and $\widetilde{\mathbf{S}}_{jj}$, are irreducible by coupled unitary transformations, as follows from Section III. The solutions are given by the following lemma, whose proof can be found in [23] (the lemma in [23] is more general, here we adapt it for the specific data in hand).

Lemma C.1. Let $\widetilde{\mathbf{S}}_{jj} \in \mathbb{R}^{Km_j \times Km_j}$ and $\widetilde{\mathbf{S}}_{ii} \in \mathbb{R}^{Km_i \times Km_i}$, $i \neq j$, $K \geq 2$, be two symmetric matrices, irreducible by coupled unitary transformations, whose (k,l)th blocks, $\widetilde{\mathbf{S}}_{jj}^{[k,l]} \in \mathbb{R}^{m_j \times m_j}$ and $\widetilde{\mathbf{S}}_{ii}^{[k,l]} \in \mathbb{R}^{m_i \times m_i}$, $k,l = 1,\ldots,K$, are nonsingular matrices. Let $\mathbf{L}^{[k]} \in \mathbb{R}^{m_i \times m_j}$ be fixed matrices such that (59) holds. Then either $\mathbf{L}^{[k]} = \mathbf{0}_{m_i \times m_j}$ $\forall k$, or $\mathbf{L}^{[k]} = \nu \mathbf{O}^{[k]}$ $\forall k$ with $\nu \in \mathbb{R}$ and $\mathbf{O}^{[k]}$ orthogonal $\forall k$ (implicitly, $m_i = m_j$).

By Lemma C.1, a non-trivial solution to (59) exists iff

$$\widetilde{\mathbf{S}}_{jj}^{[k,l]} = \mathbf{O}^{[k]}\widetilde{\mathbf{S}}_{ii}^{[k,l]}\mathbf{O}^{-[l]} \quad \forall k, l\,, \tag{60}$$

which is possible only if $m_i = m_j\ \forall k$. It remains to restate Lemma C.1 in terms of the unnormalized covariances $\mathbf{S}^{[k,l]}$. Substituting (57) in (60),

$$\mathbf{S}_{jj}^{[k,l]} = ((\mathbf{S}_{jj}^{[k,k]})^{\frac{1}{2}}\mathbf{O}^{-[k]}(\mathbf{S}_{ii}^{[k,k]})^{-\frac{1}{2}})\cdot \\ \cdot\, \mathbf{S}_{ii}^{[k,l]}((\mathbf{S}_{ii}^{[l,l]})^{-\frac{1}{2}\top}\mathbf{O}^{[l]}(\mathbf{S}_{jj}^{[l,l]})^{\frac{1}{2}\top})\,, \tag{61}$$

and setting

$$\mathbf{\Psi}^{[k]} \triangleq (\mathbf{S}_{jj}^{[k,k]})^{\frac{1}{2}}\mathbf{O}^{-[k]}(\mathbf{S}_{ii}^{[k,k]})^{-\frac{1}{2}}\,, \tag{62}$$

we obtain (14). Finally, Proposition III.2 guarantees that irreducibility by coupled unitary transformation of $\widetilde{\mathbf{S}}_{jj}$ and $\widetilde{\mathbf{S}}_{ii}$ (which is a prerequisite in Lemma C.1) is a necessary and sufficient condition for their unnormalized forms $\mathbf{S}_{jj}$ and $\mathbf{S}_{ii}$ to be irreducible by coupled congruence. This concludes the proof of Theorem V.1.

References

[1] D. Lahat and C. Jutten, "Joint independent subspace analysis using second-order statistics," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4891–4904, Sep. 2016.

[2] ——, "Joint blind source separation of multidimensional components: Model and algorithm," in *Proc. EUSIPCO*, Lisbon, Portugal, Sep. 2014, pp. 1417–1421.

[3] 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.

[4] P. Comon, "Supervised classification, a probabilistic approach," in *Proc. ESANN*, Brussels, Belgium, Apr. 1995, pp. 111–128.

[5] J.-F. Cardoso, "Multidimensional independent component analysis," in *Proc. ICASSP*, vol. 4, Seattle, WA, May 1998, pp. 1941–1944.

[6] T. Kim, T. Eltoft, and T.-W. Lee, "Independent vector analysis: An extension of ICA to multivariate components," in *Independent Component Analysis and Blind Signal Separation*, ser. LNCS, vol. 3889. Springer Berlin Heidelberg, 2006, pp. 165–172.

[7] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[8] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[9] P. Comon and C. Jutten, Eds., *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, 1st ed. Academic Press, Feb. 2010.

[10] D. Lahat and C. Jutten, "Joint analysis of multiple datasets by cross-cumulant tensor (block) diagonalization," in *Proc. SAM*, Rio de Janeiro, Brazil, Jul. 2016.

[11] ——, "Joint independent subspace analysis: A quasi-Newton algorithm," in *Latent Variable Analysis and Signal Separation*, ser. LNCS, vol. 9237. Springer International Publishing Switzerland, 2015, pp. 111–118.

[12] 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.

[13] 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.

[14] 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.

[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] X.-F. Gong, Q.-H. Lin, F.-Y. Cong, and L. De Lathauwer, "Double coupled canonical polyadic decomposition for joint blind source separation," *arXiv:1612.09466 [stat.ML]*, 2017.

[17] J. Vía, M. Anderson, X.-L. Li, and T. Adalı, "Joint blind source separation from second-order statistics: Necessary and sufficient identifiability conditions," in *Proc. ICASSP*, Prague, Czech Republic, May 2011, pp. 2520–2523.

[18] D. Lahat and C. Jutten, "An alternative proof for the identifiability of independent vector analysis using second order statistics," in *Proc. ICASSP*, Shanghai, China, Mar. 2016.

[19] D. Lahat, J.-F. Cardoso, and H. Messer, "Identifiability of second-order multidimensional ICA," in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 1875–1879.

[20] D. Lahat and C. Jutten, "A new link between joint blind source separation using second order statistics and the canonical polyadic decomposition," in *Latent Variable Analysis and Signal Separation*, ser. LNCS, Y. Deville, S. Gannot, R. Mason, M. D. Plumbley, and D. Ward, Eds. Springer, 2018.

[21] ——, "Joint independent subspace analysis by coupled block decomposition: Non-identifiable cases," in *Proc. ICASSP*, Calgary, Canada, Apr. 2018.

[22] T. J. Rothenberg, "Identification in parametric models," *Econometrica*, vol. 39, no. 3, pp. 577–591, May 1971.

[23] D. Lahat, C. Jutten, and H. Shapiro, "Schur's lemma for coupled reducibility and coupled normality," 2018, submitted.

[24] D. Lahat and C. Jutten, "A generalization to Schur's lemma with an application to joint independent subspace analysis," GIPSA-Lab, Grenoble, France, Tech. Rep. hal-01247899, Dec. 2015. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01247899

[25] H. Bousbia-Salah, A. Belouchrani, and K. Abed-Meraim, "Blind separation of non stationary sources using joint block diagonalization," in *Proc. SSP*, Aug. 2001, pp. 448–451.

[26] 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.

[27] H. W. Gutch and F. J. Theis, "Uniqueness of linear factorizations into independent subspaces," *J. Multivar. Anal.*, vol. 112, pp. 48–62, Nov. 2012.

[28] I. Schur, "Neue Begründung der Theorie der Gruppencharaktere," in *Sitzungsberichte der Königlich-Preussischen Akademie der Wissenschaften zu Berlin*, Berlin, Germany, Jan.–Jun. 1905, pp. 406–432. [Online]. Available: https://archive.org/details/sitzungsberichte1905deutsch

[29] 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.

[30] D. Lahat, J.-F. Cardoso, and H. Messer, "Blind separation of multidimensional components via subspace decomposition: Performance analysis," *IEEE Trans. Signal Process.*, vol. 62, no. 11, pp. 2894–2905, Jun. 2014.

[31] S.-J. Kim, V. D. Calhoun, and T. Adalı, "Flexible large-scale fMRI analysis: A survey," in *Proc. ICASSP*, New Orleans, LA, USA, Mar. 2017, pp. 6319–6323.

[32] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by source subspace separation," in *Proc. IEEE SP/ATHOS Workshop on HOS*, Girona, Spain, Jun. 1995, pp. 134–138.

[33] J. Ylipaavalniemi and R. Vigário, "Subspaces of spatially varying independent components in fMRI," in *Independent Component Analysis and Signal Separation*, ser. LNCS, M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, Eds., vol. 4666. Springer, 2007.

[34] B. Hunyadi, D. Camps, L. Sorber, W. Van Paesschen, M. De Vos, S. Van Huffel, and L. De Lathauwer, "Block term decomposition for modelling epileptic seizures," *EURASIP JASP*, vol. 2014, no. 14-04, p. 139, 2014.

[35] J.-F. Cardoso, M. Le Jeune, J. Delabrouille, M. Betoule, and G. Patanchon, "Component separation with flexible models – application to multichannel astrophysical observations," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 5, pp. 735–746, Oct. 2008.

[36] M. Remazeilles, J. Delabrouille, and J.-F. Cardoso, "Foreground component separation with generalized Internal Linear Combination," *MNRAS*, vol. 418, no. 1, pp. 467–476, 2011.

[37] M. A. Casey and A. Westner, "Separation of mixed audio sources by independent subspace analysis," in *Proc. ICMC*, Berlin, Germany, 8 2000, pp. 154–161.

[38] X.-L. Li, T. Adalı, and M. Anderson, "Joint blind source separation by generalized joint diagonalization of cumulant matrices," *Signal Process.*, vol. 91, no. 10, pp. 2314–2322, Oct. 2011.

[39] S. Gobillot, J. Fontecave-Jallon, V. Equy, B. Rivet, P. Gumery, and P. Hoffmann, "Non-invasive fetal monitoring using electrocardiography and phonocardiography: A preliminary study," *J. Gynecol. Obstet. Hum. Reprod.*, 2018.

[40] E. Acar, G. Gürdeniz, F. Savorani, L. Hansen, A. Olsen, A. Tjønneland, L. O. Dragsted, and R. Bro, "Forecasting chronic diseases using data fusion," *J. Proteome Res.*, vol. 16, no. 7, pp. 2435–2444, 2017.

[41] Y. Levin-Schwartz, V. D. Calhoun, and T. Adalı, "Quantifying the interaction and contribution of multiple datasets in fusion: Application to the detection of schizophrenia," *IEEE Trans. Med. Imag.*, vol. 36, no. 7, Jul. 2017.

[42] 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.

[43] D. Lahat, "A data fusion perspective on source separation," Talk presented at the *Hausdorff School: Low-rank Tensor Techniques in Numerical Analysis and Optimization*, Bonn, Germany, Apr. 2016.

[44] T. Adalı, D. Lahat, and C. Jutten, "Data fusion: Benefits of fully exploiting diversity," Tutorial presented at *EUSIPCO*, Budapest, Hungary, Aug. 2016.

[45] R. A. Horn and R. Mathias, "Block-matrix generalizations of Schur's basic theorems on Hadamard products," *Linear Algebra and its Applications*, vol. 172, pp. 337–346, Jul. 1992.

[46] 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.

[47] D. Lahat, J.-F. Cardoso, and H. Messer, "Second-order multidimensional ICA: Performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4598–4610, Sep. 2012.

[48] Y.-O. Li, T. Adalı, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3918–3929, Oct. 2009.

[49] T. W. Anderson, *An introduction to multivariate statistical analysis*. John Wiley & Sons, 1958.

[50] J. Vía, M. Anderson, X.-L. Li, and T. Adalı, "A maximum likelihood approach for independent vector analysis of Gaussian data sets," in *Proc. MLSP*, Beijing, China, Sep. 2011.

[51] L. Tong, V. C. Soon, Y. F. Huang, and R. Liu, "AMUSE: a new blind identification algorithm," in *Proc. ISCAS*, vol. 3, May 1990, pp. 1784–1787.

[52] D. Lahat, "Second-order multidimensional independent component analysis: Theory and methods," Ph.D. dissertation, Tel Aviv University, Tel Aviv, Israel, Jan. 2013. [Online]. Available: http://theses.eurasip.org/theses/640/second-order-multidimensional-independent/download/

[53] J. B. Kruskal, "Three-way arrays: rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics," *Proc. London Math. Soc.*, vol. 18, no. 2, pp. 95–138, 1977.

[54] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 661–672, Mar. 1996.

[55] H. L. Van Trees, *Optimum array processing: Part IV of detection, estimation, and modulation theory*. John Wiley & Sons, 2004.

[56] J. R. Magnus and H. Neudecker, "The commutation matrix: Some properties and applications," *Ann. Statist.*, vol. 7, no. 2, pp. 381–394, Mar. 1979.

[57] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, ser. Mathematics and its Applications. Chichester, West Sussex, England: Ellis Horwood Limited, 1981.

[58] 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.

[59] Z. Al Zhour and A. Kiliçman, "Some new connections between matrix products for partitioned and non-partitioned matrices," *Computers & Mathematics with Applications*, vol. 54, no. 6, pp. 763–784, Sep. 2007.

[60] 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.

[61] 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.

[62] C. R. Rao and S. K. Mitra, *Generalized inverse of matrices and its applications*. John Wiley & Sons, 1971, vol. 7.

[63] 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.

[64] L. R. Tucker, "The extension of factor analysis to three-dimensional matrices," in *Contributions to Mathematical Psychology*, H. Gulliksen and N. Frederiksen, Eds. New York: Holt, Rinehardt & Winston, 1964, pp. 109–127.