



HAL
open science

Joint Independent Subspace Analysis Using Second-Order Statistics

Dana Lahat, Christian Jutten

► **To cite this version:**

Dana Lahat, Christian Jutten. Joint Independent Subspace Analysis Using Second-Order Statistics. 2015. hal-01132297v1

HAL Id: hal-01132297

<https://hal.science/hal-01132297v1>

Preprint submitted on 16 Mar 2015 (v1), last revised 2 Aug 2016 (v5)

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

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

Joint Independent Subspace Analysis Using Second-Order Statistics

Dana Lahat and Christian Jutten, *Fellow, IEEE*

Abstract—This paper deals with a novel generalization of classical blind source separation (BSS)/independent component analysis (ICA) in two directions. First, we relax the constraint that the latent sources must be statistically independent. This generalization is well-known and often termed independent subspace analysis (ISA). Second, we deal with joint analysis of several such ISA problems, where the link between mixtures is formed by statistical dependence across corresponding sources in different mixtures. For the case that the data is *one-dimensional*, i.e., multiple ICA problems, this model, termed independent vector analysis (IVA), is well-known and has already been studied. Therefore, in this work, we generalize IVA to multidimensional components and term this new model joint ISA (JISA). We provide full performance analysis of this new model, including closed-form expressions for minimal mean square error (MSE), Fisher information matrix (FIM) and Cramér-Rao lower bound (CRLB) in the separation of Gaussian data. We prove in theory and validate in numerical simulations that this analysis predicts the MSE also for non-Gaussian data, when second-order statistics (SOS) are used. We present a Newton-based algorithm that converges in a significantly smaller number of iterations than the previously proposed relative gradient (RG) approach. We show that all our results indeed generalize previously-known results on IVA via SOS, including the ability to resolve static mixtures of Gaussian stationary data and the individual arbitrary permutation. Finally, we discuss some links between this model and BSS of non-stationary multidimensional data.

Index Terms—Joint blind source separation, independent vector analysis, independent subspace analysis, multidimensional independent component analysis, joint block diagonalization, second-order methods, performance analysis, algorithms

I. INTRODUCTION

In this work, we present a model inspired by two recently-proposed extensions to blind source separation (BSS) that until now have been dealt with only separately: (1) relaxing the constraint that latent sources within a set of measurements must be statistically independent, often termed multidimensional independent component analysis (MICA) or independent subspace analysis (ISA) [1]–[3], and (2) solving several BSS problems simultaneously by exploiting statistical dependencies between latent sources across sets of measurements, a model usually known as independent vector analysis (IVA) or joint BSS (JBSS) [4], [5]. The new model, termed joint ISA (JISA) [6], is a generalization of JBSS to multidimensional components.

The idea of solving the multidimensional components problem in terms of subspace separation through independent

component analysis (ICA) was first demonstrated in [1], on fetal electrocardiography (ECG) recordings. The perspective of *multidimensional* ICA, of vector-valued components whose representation is based on unambiguous projections on the sources’ respective subspaces, was presented in [1], and an elaborate geometric framework was suggested in [2]. A prevalent approach for ISA consists of using ICA-based algorithms followed by a clustering step [7]–[10]. Algorithms that exploit the true multidimensional nature of the data can be found, for example, in [3], [11]–[16]. A theoretical analysis of the advantage, in terms of component estimation error, of using the true multidimensional model over the more prevalent two-step approach of BSS followed by a clustering step is given in [17] for real Gaussian piecewise-stationary data. Identifiability and uniqueness of decompositions into invariant subspaces of dimensions larger than one are discussed in [18]–[21].

Multidimensional data may occur due to various complex relations and processes within the underlying phenomena. The dimension of a dependent group may not always reflect the number of its underlying physical elements. As a result, in multidimensional models, there is not always a physically meaningful interpretation to separating the multidimensional components back into single-dimensional elements. This holds, for example, in neurological activity observed by functional magnetic resonance imaging (fMRI) [22] and electroencephalography (EEG) [23], fetal ECG [1], [2], natural images [3] and astrophysical processes [24]. For such data, a one-dimensional model is often just an approximation. In this work, we focus on separation into subspaces that represent statistically independent multivariate components. Further decomposition, within a dependent group, if admissible by the application, is beyond the scope of this paper.

One of the earliest frameworks to simultaneously analyse several datasets through statistical links between their latent parameters is canonical correlation analysis (CCA) [25]. The idea to simultaneously solve several ICA problems by exploiting higher-order statistical *dependence* between latent sources *across sets of measurements* was introduced by Kim et al. [4], [26], and termed IVA. The method has been shown to be able to resolve the permutation ambiguity that is inherent to classical ICA up to a *single* permutation matrix that is common to all sets of measurements [4], [26]. Li et al. [5] have shown that the IVA framework, which they termed JBSS, provides sufficient constraints for identifying real Gaussian stationary processes that had been mixed by an invertible matrix, a problem that is ill-posed with classical BSS/ICA, where each mixture is processed separately [27]. Li et al. [28] have shown that JBSS can be formulated as a coupled matrix diagonaliza-

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

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

tion problem that minimizes a quadratic criterion, and solved by exploiting either second- or higher-order statistics. This observation, that coupled matrix factorizations enjoy more relaxed uniqueness conditions, finds its tensor counterpart in [21]. Recently, JBSS algorithms that minimize the maximum likelihood (ML), mutual information (MI) and entropy have been proposed [29], [30]. When only two datasets are involved, second-order statistics (SOS) JBSS amounts to CCA and can be solved in closed-form using generalized eigenvalue decomposition (GEVD) [31, Chapter 12]. A comprehensive theoretical analysis of IVA can be found in [30] and references therein.

Considering the growing evidence of IVA as a helpful tool in various applications such as multiset data analysis [5], [28], [30], hyperscanning [32] and dynamic systems [33], and the fact that natural signals are often better modelled as multidimensional, it is only natural to take advantage of the benefits of both.

The JISA model, which is the core of this paper, and a SOS-based relative gradient (RG) algorithm that achieves the optimal separation in the presence of Gaussian data, were first presented in [6]. The novelty and contribution of this paper is in providing a comprehensive theoretical analysis to the SOS approach in [6], including closed-form expressions for the mean square error (MSE), Fisher information matrix (FIM) and Cramér-Rao lower bound (CRLB), that also results in a quasi-Newton (QN) algorithm that convergence in a significantly smaller number of iterations than the previously-proposed RG approach. We mention that a gradient algorithm that performs JISA based on the multivariate Laplace distribution has recently been proposed in [34].

In this paper, we adopt the analytical approach based on small-error analysis that was used in [17], [35] to analyse the performance of non-stationary multidimensional BSS. Although the two models are essentially different, this approach makes some interesting similarities and analogies between the two models be manifested.

The following notations and conventions are used throughout this paper. Bold lowercase letters denote vectors; regular lowercase letters denote scalars. Bold uppercase letters denote matrices; regular uppercase letters denote functions or operators; calligraphic uppercase letters denote sets. For simplicity, we assume that all values are real. Trace is denoted by $\text{tr}\{\cdot\}$; $(\cdot)^\dagger$ denotes transpose. $|\mathbf{a}|^2 = \mathbf{a}^\dagger \mathbf{a}$ for any vector \mathbf{a} . $\mathbf{A}^{-\dagger} = (\mathbf{A}^{-1})^\dagger$ whenever the inverse exists. $\text{vec}\{\cdot\}$ denotes the operator that stacks the columns of a $P \times Q$ matrix into a $PQ \times 1$ vector. The direct sum of K rectangular matrices $\mathbf{M}^{[k]}$ is denoted by $\bigoplus_{k=1}^K \mathbf{M}^{[k]}$ and yields a block-diagonal matrix with $\mathbf{M}^{[k]}$ as its k th diagonal element. The operator $\text{bdiag}_{\mathbf{b}}\{\mathbf{M}\}$, given a $P \times P$ matrix \mathbf{M} and a vector \mathbf{b} of positive integers that sum up to P , extracts from \mathbf{M} a block-diagonal matrix with block-pattern \mathbf{b} and zeroes the off-diagonal blocks. $\text{bdiag}_{\mathbf{b}}^{-1}\{\cdot\}$ stands for $(\text{bdiag}_{\mathbf{b}}\{\cdot\})^{-1}$. $\mathcal{B}_{\mathbf{b}}$ denotes the subspace of all invertible block-diagonal matrices with block-pattern \mathbf{b} . $\mathbf{0}$ denotes a one- or two-dimensional array of zeros. $\mathbf{1}_P$ denotes an $P \times 1$ vector of ones. \mathbf{I}_P stands for the $P \times P$ identity matrix, with dimensions that

are omitted if they are implicit. $E\{\cdot\}$ denotes expectation. $\text{Cov}(\mathbf{a}) = E\{\mathbf{a}\mathbf{a}^\dagger\}$, $\text{Cov}(\mathbf{a}, \mathbf{b}) = E\{\mathbf{a}\mathbf{b}^\dagger\}$ for any stochastic vectors \mathbf{a}, \mathbf{b} with $E\{\mathbf{a}\} = \mathbf{0}$. $\|\cdot\|^2$ denotes the Frobenius norm; δ_{ij} denotes the Kronecker delta. The Kronecker product is denoted by \otimes . 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. Then, the Khatri-Rao product [36]–[39] is defined as $\mathbf{A} \odot \mathbf{B} = \begin{bmatrix} \mathbf{A}_{11} \otimes \mathbf{B}_{11} & \mathbf{A}_{12} \otimes \mathbf{B}_{12} & \cdots \\ \mathbf{A}_{21} \otimes \mathbf{B}_{21} & \mathbf{A}_{22} \otimes \mathbf{B}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix}$, where $\mathbf{A}_{ij} \otimes \mathbf{B}_{ij}$ and $\mathbf{A} \odot \mathbf{B}$ are $m_i p_i \times n_j q_j$ and $(\sum m_i p_i) \times (\sum n_j q_j)$, respectively. $\Omega(f)$ stands for zero-mean stochastic terms whose standard deviation is proportional to f , or to higher powers thereof. $O(f)$ stands for deterministic terms that are bounded above, up to a constant factor, by f .

The rest of this paper is organized as follows. In Sec. II we present and define the model that we denote JISA, and formalize it mathematically. In Sec. III we present a SOS approach to JISA. Sec. II–III recall results from [6], whereas the novelty is in the sections that follow. Sec. IV provides a theoretical SOS error analysis of the proposed approach. Sec. V briefly discusses the well-posedness of the model. In Sec. VI we present a QN algorithm that asymptotically achieves the theoretical error derived in Sec. IV and thus the minimal mean square error (MMSE) for Gaussian data. Numerical experiments in Sec. VII validate the our results. We conclude our paper with a discussion in Sec. VIII.

II. JISA: MODEL AND PROBLEM FORMULATION

Consider T observations of K vectors $\mathbf{x}^{[k]}(t)$, modelled as

$$\mathbf{x}^{[k]}(t) = \mathbf{A}^{[k]} \mathbf{s}^{[k]}(t) \quad 1 \leq t \leq T, 1 \leq k \leq K, \quad (1)$$

where $\mathbf{A}^{[k]}$ are $M \times M$ invertible matrices that may be different $\forall k$, and $\mathbf{x}^{[k]}(t)$ and $\mathbf{s}^{[k]}(t)$ are $M \times 1$ vectors. For fixed k , each mixture (1) corresponds to classical BSS. In JBSS, the elements of the $K \times 1$ vector $\mathbf{s}_i^{\text{JBSS}}(t) = [s_i^{[1]}(t), \dots, s_i^{[K]}(t)]^\dagger$, $i = 1, \dots, M$, are statistically *dependent* whereas the pairs $(\mathbf{s}_i^{\text{JBSS}}(t), \mathbf{s}_j^{\text{JBSS}}(t))$ are statistically *independent* for all $i \neq j \in \{1, \dots, M\}$. Therefore, JBSS aims at extracting M mutually independent vector elements (whence the name IVA) from K sets of measurements by exploiting not only the statistical *independence within each set of measurements* but also the *dependence across sets of measurements*.

Given the partition $\mathbf{s}^{[k]}(t) = [\mathbf{s}_1^{[k]\dagger}(t), \dots, \mathbf{s}_N^{[k]\dagger}(t)]^\dagger$, where $N \leq M$, $\mathbf{s}_i^{[k]}(t)$ are $m_i \times 1$ vectors, $m_i \geq 1$, $\sum_{i=1}^N m_i = M$, and the probability density function (pdf) of each m_i -dimensional random process $\mathbf{s}_i^{[k]}(t)$ irreducible in the sense that it cannot be factorized into a product of non-trivial pdfs, then each mixture (1) represents a single ISA problem. The *model* that we define¹ as JISA corresponds to linking several such ISA problems via the assumption that the elements of

¹This formulation is sufficiently simple to keep notations and derivations easily tractable yet at the same time sufficiently rich to encompass the essential properties of JISA. Our results generalize straightforwardly to multidimensional components of different dimensions, as long as $\mathbf{A}^{[k]}$, now of different dimensions, remain invertible. One can also relax the statistical dependence across mixtures, as long as uniqueness is preserved, see also Sec. V, footnote 4 and Sec. VIII.

the $n_i \times 1$ vector $\mathbf{s}_i(t) = [\mathbf{s}_i^{[1]\dagger}(t), \dots, \mathbf{s}_i^{[K]\dagger}(t)]^\dagger$, where $n_i = Km_i$, are statistically *dependent* whereas the pairs $(\mathbf{s}_i(t), \mathbf{s}_j(t))$ are statistically *independent* for all $i \neq j \in \{1, \dots, N\}$.

Given $\mathcal{M} = \{m_i\}_{i=1}^N$ and the set of observations $\mathcal{X} = \{\mathbf{x}^{[k]}(t)\}_{k=1, t=1}^{K, T}$, the *problem* of JISA is that of finding linear transformations $\mathbf{A}^{-[k]}$ on \mathcal{X} such that the source vectors $\mathbf{s}_1(t), \dots, \mathbf{s}_N(t)$ are as independent as possible. This notion is given a definite meaning in Section III where we set up a simple statistical model that, via its likelihood function, yields a quantitative measure of independence.

The above partition of $\mathbf{s}^{[k]}(t)$ induces a corresponding partition in the mixing matrices: $\mathbf{A}^{[k]} = [\mathbf{A}_1^{[k]}] \cdots [\mathbf{A}_N^{[k]}]$ with $\mathbf{A}_i^{[k]}$ the i th $M \times m_i$ column-block of $\mathbf{A}^{[k]}$. The multiplicative model (1) may now be rewritten as a sum of $N \leq M$ *multidimensional components*:

$$\mathbf{x}^{[k]}(t) = \sum_{i=1}^N \mathbf{x}_i^{[k]}(t) \quad (2)$$

where the i th $M \times 1$ component $\mathbf{x}_i^{[k]}(t)$ is defined as

$$\mathbf{x}_i^{[k]}(t) = \mathbf{A}_i^{[k]} \mathbf{s}_i^{[k]}(t). \quad (3)$$

In a blind context, the component vector $\mathbf{x}_i^{[k]}(t)$ is better defined than the source vector $\mathbf{s}_i^{[k]}(t)$. Indeed, for any invertible $m_i \times m_i$ matrix $\mathbf{Z}_{ii}^{[k]}$, it is impossible to discriminate between the representation of a component $\mathbf{x}_i^{[k]}(t)$ by the pair $(\mathbf{A}_i^{[k]}, \mathbf{s}_i^{[k]}(t))$ and $(\mathbf{A}_i^{[k]} \mathbf{Z}_{ii}^{-[k]}, \mathbf{Z}_{ii}^{[k]} \mathbf{s}_i^{[k]}(t))$, where $\mathbf{Z}_{ii}^{-[k]}$ denotes $(\mathbf{Z}_{ii}^{[k]})^{-1}$. This means that only the column space of $\mathbf{A}_i^{[k]}$, $\text{span}(\mathbf{A}_i^{[k]})$, can be blindly identified. Therefore, JISA is in fact a (joint) subspace estimation problem. We further note that (2) is invariant to arbitrary permutation of the source indices. If all the linear transformations on $\{\mathbf{x}^{[k]}(t)\}$ that maximize statistical independence between $\mathbf{s}_1(t), \dots, \mathbf{s}_N(t)$ yield the same components, we say that the JISA model is unique.

Further insights can be obtained by rewriting (1) as

$$\mathbf{x}(t) = \mathbf{A} \mathbf{s}(t) \quad (4)$$

where $\mathbf{s}(t) = [\mathbf{s}^{[1]\dagger}(t), \dots, \mathbf{s}^{[K]\dagger}(t)]^\dagger$ and $\mathbf{x}(t) = [\mathbf{x}^{[1]\dagger}(t), \dots, \mathbf{x}^{[K]\dagger}(t)]^\dagger$ are $L \times 1$ vectors, where $L = KM$, and $\mathbf{A} = \bigoplus_{k=1}^K \mathbf{A}^{[k]} \in \mathcal{B}_k$ where $\mathbf{k} = M\mathbf{1}_K$. With these notations, $\tilde{\mathbf{s}}(t) = \Phi \mathbf{s}(t)$, where $\tilde{\mathbf{s}}(t) = [\mathbf{s}_1^\dagger(t), \dots, \mathbf{s}_N^\dagger(t)]^\dagger$ and Φ is the corresponding $L \times L$ permutation matrix. Combining (2) and (3) in (4), one obtains that $\mathbf{x}(t) = \sum_{i=1}^N (\mathbf{I}_K \odot \mathbf{A}_i) \mathbf{s}_i(t)$, where $\mathbf{A}_i \triangleq [\mathbf{A}_i^{[1]}] \cdots [\mathbf{A}_i^{[K]}]$ and \mathbf{I}_K partitioned columnwise. The latter implies that $\mathbf{x}(t)$ is invariant to permutation of the indices $i = 1, \dots, N$ for any arbitrary ordering of $\{1, \dots, K\}$ (the ordering of the mixtures is chosen by the user and thus known). This simple observation amounts to the fact that the indexing of the latent components must be the same in all the mixtures. This proves that the ability of IVA to provide a single permutation matrix to all the involved mixtures (Sec. I) indeed generalizes to the multidimensional case.

As we shall see later, it is useful to introduce the separating projectors: these are the $M \times M$ oblique projection matrices $\mathbf{P}_i^{[k]}$ onto $\text{span}(\mathbf{A}_i^{[k]})$ along $\text{span}(\mathbf{A}_j^{[k]}) \forall j \neq i$. By definition,

they satisfy $\mathbf{P}_i^{[k]} \mathbf{A}_j^{[k]} = \delta_{ij} \mathbf{A}_i^{[k]}$, unaffected if $\mathbf{A}_i^{[k]}$ is changed into $\mathbf{A}_i^{[k]} \mathbf{Z}_{ii}^{-[k]}$ and, most importantly, allow one to write

$$\mathbf{x}_i^{[k]}(t) = \mathbf{P}_i^{[k]} \mathbf{x}^{[k]}(t). \quad (5)$$

Finally, note that if $\mathbf{B}^{[k]} = \mathbf{A}^{-[k]}$ is partitioned into N horizontal $m_i \times M$ blocks $\mathbf{B}_i^{[k]}$, then the rank- m_i i th oblique projection is given by

$$\mathbf{P}_i^{[k]} = \mathbf{A}_i^{[k]} \mathbf{B}_i^{[k]}. \quad (6)$$

Alternatively, one can define oblique projections such that $\mathbf{P}_i \mathbf{x}(t) = \mathbf{x}_i(t)$, $\mathbf{x}_i(t) = [\mathbf{x}_i^{[1]\dagger}(t), \dots, \mathbf{x}_i^{[K]\dagger}(t)]^\dagger$. It is easy to verify that $\mathbf{P}_i = \bigoplus_{k=1}^K \mathbf{P}_i^{[k]}$.

In the rest of this paper, we focus on JISA using SOS. Assuming sample independence $\forall t \neq t'$, the model assumptions imply that $\tilde{\mathbf{S}} \triangleq E\{\tilde{\mathbf{s}}(t) \tilde{\mathbf{s}}^\dagger(t)\} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{NN} \end{bmatrix} = \bigoplus_{i=1}^N \mathbf{S}_{ii} \in \mathcal{B}_n$, where $\tilde{\mathbf{S}}$ is a $L \times L$ block-diagonal matrix with block-pattern $\mathbf{n} = [n_1, \dots, n_N]^\dagger$ and

$$\tilde{\mathbf{S}} = \Phi \mathbf{S} \Phi^\dagger \in \mathcal{B}_n. \quad (7)$$

The (i, j) th $n_i \times n_j$ block of $\tilde{\mathbf{S}}$ is $\mathbf{S}_{ij} = E\{\mathbf{s}_i(t) \mathbf{s}_j^\dagger(t)\}$ for $1 \leq i, j \leq N$. Its empirical counterpart is $\tilde{\mathbf{S}}_{ij} = \frac{1}{T} \sum_{t=1}^T \mathbf{s}_i(t) \mathbf{s}_j^\dagger(t)$. The linear model (4) implies that $\mathbf{X} = \mathbf{A} \mathbf{S} \mathbf{A}^\dagger$ where $\mathbf{S} = E\{\mathbf{s}(t) \mathbf{s}^\dagger(t)\}$ and $\mathbf{X} = E\{\mathbf{x}(t) \mathbf{x}^\dagger(t)\}$. In the sequel, we assume that all \mathbf{S}_{ii} are invertible and do not contain values fixed to zero; in practice, this assumption could be relaxed, as explained in footnote 1; further discussion of this matter is beyond the scope of this work. Typical structures of \mathbf{A} , Φ , \mathbf{S} , $\tilde{\mathbf{S}}$ and \mathbf{X} are illustrated in Fig. 1.

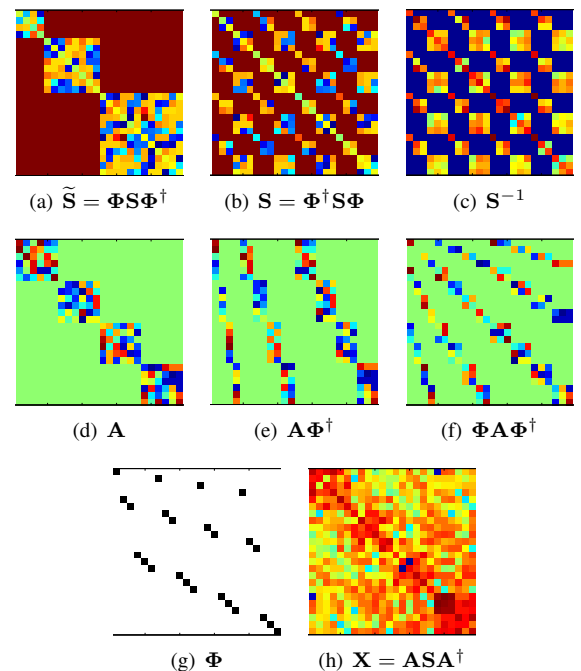


Fig. 1. Structure of typical matrices used in JISA. $\mathcal{M} = \{1, 2, 3\}$, $K = 4$. False colours.

III. OPTIMAL COMPONENT SEPARATION USING SOS

In the following, we consider a Gaussian model in which $\mathbf{s}_i(t) \sim \mathcal{N}(\mathbf{0}_{n_i \times 1}, \mathbf{S}_{ii})$ are mutually independent samples $\forall t \neq t'$. The log-likelihood for the model just described is

$$\begin{aligned} \log p(\mathcal{X}; \mathcal{A}, \mathbf{S}) &\triangleq -T\phi(\mathcal{A}, \mathbf{S}) = \sum_{t=1}^T \log p(\mathbf{x}(t)) \\ &= -\frac{1}{2} \sum_{t=1}^T (\log \det 2\pi \mathbf{X} + \mathbf{x}^\dagger(t) \mathbf{X}^{-1} \mathbf{x}(t)) \\ &= -\frac{T}{2} (\log \det 2\pi \mathbf{X} + \text{tr}\{\bar{\mathbf{X}} \mathbf{X}^{-1}\}) \\ &= -TD(\bar{\mathbf{X}}, \mathbf{X}) - \kappa = -TD(\bar{\mathbf{X}}, \mathbf{A} \mathbf{S} \mathbf{A}^\dagger) - \kappa \\ &= -TD(\Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger} \Phi^\dagger, \tilde{\mathbf{S}}) - \kappa \end{aligned} \quad (8)$$

where $\mathcal{A} = \{\mathbf{A}^{[k]}\}_{k=1}^K$ and $\bar{\mathbf{X}} = \frac{1}{T} \sum_{t=1}^T \mathbf{x}(t) \mathbf{x}^\dagger(t)$ is the empirical counterpart of \mathbf{X} . The term $\kappa = \frac{T}{2} (\log \det(2\pi \bar{\mathbf{X}}) + L)$ is irrelevant to the maximization of the likelihood since it depends only on the data and not on the parameters. The second equality in (8) is due to the assumption of pairwise sample independence for $t \neq t'$. The third equality is due to the Gaussian assumption and (4) that imply $\mathbf{x}(t) \sim \mathcal{N}(\mathbf{0}_{L \times 1}, \mathbf{X})$. The fourth equality uses the property $\mathbf{a}^\dagger \mathbf{R} \mathbf{a} = \text{tr}\{\mathbf{R} \mathbf{a} \mathbf{a}^\dagger\}$ for any vector \mathbf{a} and matrix \mathbf{R} of appropriate dimensions. The scalar

$$D(\mathbf{R}_1, \mathbf{R}_2) = \frac{1}{2} (\text{tr}\{\mathbf{R}_1 \mathbf{R}_2^{-1}\} - \log \det(\mathbf{R}_1 \mathbf{R}_2^{-1}) - M), \quad (9)$$

defined for any two $M \times M$ symmetric positive-definite matrices \mathbf{R}_1 and \mathbf{R}_2 , is the Kullback-Leibler divergence (KLD) between the distributions $\mathcal{N}(\mathbf{0}, \mathbf{R}_1)$ and $\mathcal{N}(\mathbf{0}, \mathbf{R}_2)$ [40]. The last equality in (8) follows from (7), (9) and (46d). The derivation of (8) follows similar lines as those used in [35, Sec. III] and [41, Sec. 3].

A. Maximum Likelihood with respect to (w.r.t.) the Nuisance Parameters and Contrast Function

Given the block-diagonal structure of $\tilde{\mathbf{S}}$, the last step in (8) gives rise to its ML estimate [35, Appendix B]

$$\hat{\tilde{\mathbf{S}}}^{\text{ML}} = \text{bdiag}_{\mathbf{n}}\{\Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger} \Phi^\dagger\} \quad (10a)$$

$$\hat{\mathbf{S}}^{\text{ML}} = \Phi^\dagger \text{bdiag}_{\mathbf{n}}\{\Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger} \Phi^\dagger\} \Phi \quad (10b)$$

where (10b) is due to (7). We can now write

$$\max_{\mathbf{S}} \log p(\mathcal{X}; \mathcal{A}, \mathbf{S}) = -TC(\mathbf{A}) + \kappa, \quad (11)$$

where in the latter we have defined the *contrast function* [42]

$$C(\mathbf{A}) = D(\Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger} \Phi^\dagger, \text{bdiag}_{\mathbf{n}}\{\Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger} \Phi^\dagger\}) \quad (12)$$

It holds that $D(\mathbf{R}, \text{bdiag}_{\mathbf{b}}\{\mathbf{R}\}) \geq 0$ with equality if and only if (iff) $\mathbf{R} \in \mathcal{B}_{\mathbf{b}}$. Hence, for any positive-definite matrix \mathbf{R} , $D(\mathbf{R}, \text{bdiag}_{\mathbf{b}}\{\mathbf{R}\})$ is a measure of the block-diagonality of \mathbf{R} . Therefore, minimizing the contrast function² (12) amounts to (approximate) *block diagonalization* of $\bar{\mathbf{X}}$ by a permuted block-diagonal matrix $\Phi \mathbf{A}^{-1}$.

Note 1. The result in [41, Sec. 3.3] is a special case of (10b) when $m_i = 1 \forall i$.

²We assume that an optimum exists.

B. Estimating Equations

The next step is estimating \mathcal{A} . This is achieved by characterizing the stationary points of the contrast function (12). For this purpose, we calculate the derivative of $\phi(\mathcal{A}, \mathbf{S})$ w.r.t. each $\mathbf{A}^{[k]}$ separately, for fixed \mathbf{S} and $\mathcal{A} \setminus \mathbf{A}^{[k]}$, as we now explain. Consider a relative variation $\mathbf{A}^{[k]} \rightarrow \mathbf{A}^{[k]}(\mathbf{I} + \delta^{[k]})^{-1}$, where $\delta^{[k]}$ is $M \times M$ and has arbitrary values but such that $\mathbf{I} + \delta^{[k]}$ is invertible³. Then, the first-order variation of $\phi(\mathcal{A}, \mathbf{S})$ when $\mathbf{A}^{[k]}$ is replaced by $\mathbf{A}^{[k]}(\mathbf{I} + \delta^{[k]})^{-1}$ can always be expressed by the Taylor expansion

$$\begin{aligned} \phi(\{\mathcal{A} \setminus \mathbf{A}^{[k]}, \mathbf{A}^{[k]}(\mathbf{I} + \delta^{[k]})^{-1}\}, \mathbf{S}) &= \phi(\mathcal{A}, \mathbf{S}) \\ &+ \text{tr}\{(\nabla \phi^{[k]}(\mathcal{A}, \mathbf{S}))^\dagger \delta^{[k]}\} + O(\|\delta^{[k]}\|^2), \end{aligned} \quad (13)$$

where $\nabla \phi^{[k]}(\mathcal{A}, \mathbf{S})$ denotes the $M \times M$ RG of $\phi(\mathcal{A}, \mathbf{S})$ w.r.t. $\mathbf{A}^{[k]}$. Derivation similar to [35, Sec. III.D] yields

$$\nabla \phi^{[k]}(\mathcal{A}, \mathbf{S}) = \mathbf{J}_k^\dagger \mathbf{S}^{-1} \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger} \mathbf{J}_k - \mathbf{I}_M, \quad (14)$$

where \mathbf{J}_k is the k th $L \times M$ block of $\mathbf{I}_L = [\mathbf{J}_1, \dots, \mathbf{J}_K]$. The K terms $\nabla \phi^{[k]}(\mathcal{A}, \mathbf{S})$ in (14) can be collected into

$$\begin{aligned} \nabla \phi(\mathcal{A}, \mathbf{S}) &\triangleq \sum_{k=1}^K \mathbf{J}_k \nabla \phi^{[k]}(\mathcal{A}, \mathbf{S}) \mathbf{J}_k^\dagger = \oplus_{k=1}^K \nabla \phi^{[k]}(\mathcal{A}, \mathbf{S}) \\ &= \text{bdiag}_{\mathbf{k}}\{\mathbf{S}^{-1} \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger}\} - \mathbf{I}_L. \end{aligned} \quad (15)$$

It can be shown that the first-order variation of $C(\mathbf{A})$ w.r.t. $\mathbf{A}^{[k]}$, derived similarly to (13)–(14), obeys

$$\nabla C^{[k]}(\mathbf{A}) = \nabla \phi^{[k]}(\mathcal{A}, \mathbf{S})|_{\mathbf{S}=\hat{\mathbf{S}}^{\text{ML}}}. \quad (16)$$

Given (15), (16), (10b) and $\nabla C(\mathbf{A}) = \oplus_{k=1}^K \nabla C^{[k]}(\mathbf{A})$, we can now write

$$\begin{aligned} \nabla C(\mathbf{A}) &= \nabla \phi(\mathcal{A}, \mathbf{S})|_{\mathbf{S}=\hat{\mathbf{S}}^{\text{ML}}} = \\ &\text{bdiag}_{\mathbf{k}}\{\Phi^\dagger \text{bdiag}_{\mathbf{n}}^{-1}\{\Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger} \Phi^\dagger\} \Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger}\} - \mathbf{I}. \end{aligned} \quad (17)$$

Values of \mathcal{A} that maximize the likelihood and thus minimize $C(\mathbf{A})$ also satisfy $\nabla C(\mathbf{A}) = \mathbf{0}$. Hence, in the sequel, matrices that satisfy the *estimating equations*

$$\text{bdiag}_{\mathbf{k}}\{\Phi^\dagger \text{bdiag}_{\mathbf{n}}^{-1}\{\Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger} \Phi^\dagger\} \Phi \mathbf{A}^{-1} \bar{\mathbf{X}} \mathbf{A}^{-\dagger}\} = \mathbf{I} \quad (18)$$

are denoted $\hat{\mathbf{A}}^{[k]\text{ML}}$. The associated oblique projections (6) are denoted $\hat{\mathbf{P}}_i^{[k]\text{ML}}$. The corresponding component estimates are given by

$$\hat{\mathbf{x}}_i^{[k]\text{ML}}(t) \triangleq \hat{\mathbf{P}}_i^{[k]\text{ML}} \mathbf{x}^{[k]}(t), \quad (19)$$

which follows from (5).

³It should be emphasized that \mathbf{A} is block-diagonal by definition and thus there is absolutely no meaning to perturbing its off-block-diagonal entries. This is the bifurcation point from which the derivation of the RG takes a different path than that in [35].

C. Figure of Merit: Mean Square Error

Our goal is component separation. Therefore, the problem of JISA consists in estimating $\mathbf{x}_i^{[k]}(t)$ given only \mathcal{X} and \mathcal{M} . Consequently, we define the figure of merit as the MSE in the estimation of $\mathbf{x}_i^{[k]}(t)$,

$$\widehat{\text{MSE}}_i^{[k]} = \frac{1}{T} \sum_{t=1}^T \|\widehat{\mathbf{x}}_i^{[k]}(t) - \mathbf{x}_i^{[k]}(t)\|^2, \quad (20)$$

where $\|\cdot\|$ denotes the Frobenius norm. Alternatively, we may be interested in the (normalized) MSE in the estimation of $\mathbf{x}_i(t) = [\mathbf{x}_i^{[1]\dagger}(t), \dots, \mathbf{x}_i^{[K]\dagger}(t)]^\dagger$,

$$\widehat{\text{MSE}}_i = \frac{1}{\sigma_i^2} \frac{1}{T} \sum_{t=1}^T \|\widehat{\mathbf{x}}_i(t) - \mathbf{x}_i(t)\|^2, \quad (21)$$

where $\sigma_i^2 = E\{\|\mathbf{x}_i(t)\|^2\}$. For Gaussian data, estimates of $\mathbf{x}_i(t)$ obtained via (19) from matrices that satisfy (18) achieve asymptotically (i.e., $T \rightarrow \infty$) the MMSE.

IV. ERROR ANALYSIS

We now turn to the error analysis of our model. This will lead us to a closed-form expression for the FIM and CRLB in the estimation of the oblique projections and to the MSE in component estimation.

A. Error Decomposition

A difficulty in the error analysis of blind subspace estimation stems from the inability to characterize the error in the mixing matrices, due to severe indeterminacies they suffer from (Sec. II). We thus begin by defining convenient error terms. In order to focus on well-defined quantities, we consider the errors

$$\delta \mathbf{P}_i^{[k]} \triangleq \widehat{\mathbf{P}}_i^{[k]} - \mathbf{P}_i^{[k]} \quad (22)$$

in $\widehat{\mathbf{P}}_i^{[k]}$, the estimates of the oblique projectors $\mathbf{P}_i^{[k]}$. Accordingly, the estimate of $\mathbf{x}_i^{[k]}(t)$ is $\widehat{\mathbf{x}}_i^{[k]}(t) = \widehat{\mathbf{P}}_i^{[k]} \mathbf{x}^{[k]}(t) = \mathbf{x}_i^{[k]}(t) + \delta \mathbf{P}_i^{[k]} \mathbf{x}^{[k]}(t)$, which follows from (5), (19) and (22). Consequently, the component estimation error is given by

$$\widehat{\mathbf{x}}_i^{[k]}(t) - \mathbf{x}_i^{[k]}(t) = \delta \mathbf{P}_i^{[k]} \mathbf{x}^{[k]}(t) \quad (23)$$

and (20) can be rewritten as

$$\begin{aligned} \widehat{\text{MSE}}_i^{[k]} &= \frac{1}{T} \sum_{t=1}^T \|\delta \mathbf{P}_i^{[k]} \mathbf{x}^{[k]}(t)\|^2 \\ &= \text{tr}\{(\overline{\mathbf{X}}^{[k,k]} \otimes \mathbf{I}_M) \text{vec}\{\delta \mathbf{P}_i^{[k]}\} \text{vec}^\dagger\{\delta \mathbf{P}_i^{[k]}\}\} \end{aligned} \quad (24)$$

where the last equality uses $\|\mathbf{a}\|^2 = \text{tr}\{\mathbf{a}\mathbf{a}^\dagger\}$ and Property A.1 in Appendix A. $\overline{\mathbf{X}}^{[k,l]}$ and $\mathbf{X}^{[k,l]}$ denote the (k, l) th blocks, according to block-partition \mathbf{k} , of $\overline{\mathbf{X}}$ and \mathbf{X} , respectively. It can be shown (Appendix B) that asymptotically,

$$\begin{aligned} \text{MSE}_i^{[k]} &\triangleq E\{\widehat{\text{MSE}}_i^{[k]}\} \\ &= \text{tr}\{(\mathbf{X}^{[k,k]} \otimes \mathbf{I}_M) \text{Cov}(\text{vec}\{\delta \mathbf{P}_i^{[k]}\})\} + O\left(\frac{1}{T^{3/2}}\right). \end{aligned} \quad (25)$$

In the following, we set out to obtain a closed-form expression for $\text{Cov}(\text{vec}\{\delta \mathbf{P}_i^{[k]}\})$ as a function only of the model parameters, that will conclude the derivation of the MSE.

B. First-Order Approximation of $\delta \mathbf{P}_i^{[k]}$

In general, any estimate or approximation of $\mathbf{A}^{[k]}$ can be rewritten as a product of $\mathbf{A}^{[k]}$ and some perturbation matrix. In addition, as explained in Appendix D, the contrast function (12) is invariant to right-multiplying each $\mathbf{A}^{[k]}$ by any $\mathbf{\Lambda}^{[k]} \in \mathcal{B}_m$, $\mathbf{m} = [m_1, \dots, m_N]^\dagger$. Hence, the most general form of the minimizer of (12) can be formulated as

$$\widehat{\mathbf{A}}^{[k]} = \mathbf{A}^{[k]} (\mathbf{I}_M + \mathcal{E}^{[k]})^{-1} \mathbf{\Lambda}^{[k]}, \quad (26)$$

where the $M \times M$ matrix $\mathcal{E}^{[k]}$ reflects the *relative* change in $\mathbf{A}^{[k]}$, up to the scale ambiguity which is represented by $\mathbf{\Lambda}^{[k]}$. In Appendix E we show that

$$\begin{aligned} \delta \mathbf{P}_i^{[k]} &= \mathbf{A}^{[k]} \left(\sum_{j \neq i} \mathbf{E}_i \mathcal{E}_{ij}^{[k]} \mathbf{E}_j^\dagger - \mathbf{E}_j \mathcal{E}_{ji}^{[k]} \mathbf{E}_i^\dagger \right) \mathbf{A}^{-[k]} \\ &\quad + O(\|\mathcal{E}^{[k]}\|^2) \end{aligned} \quad (27)$$

where \mathbf{E}_i is the i th $M \times m_i$ block of $\mathbf{I}_M = [\mathbf{E}_1, \dots, \mathbf{E}_N]$. The $m_i \times m_j$ matrix $\mathcal{E}_{ij}^{[k]}$ denotes the (i, j) th block of $\mathcal{E}^{[k]}$, according to partition \mathbf{m} . Since $\mathbf{\Lambda}^{[k]}$ has vanished from (27), we can proceed with our error analysis without worrying about the scale ambiguity.

C. Influence Function

In order to evaluate the covariance of the error terms, we first establish the first-order expansion of $\mathcal{E}_{ij}^{[k]}$ in terms of the sample covariance matrices. The key assumption for our blind separation is source decorrelation: $\mathbf{S}_{ij}^{[k,l]} = E\{\overline{\mathbf{S}}_{ij}^{[k,l]}\} = \mathbf{0}_{m_i \times m_j}$ for $j \neq i$ and any k, l , where

$$\overline{\mathbf{S}}_{ij}^{[k,l]} = \frac{1}{T} \sum_{t=1}^T \mathbf{s}_i^{[k]}(t) \mathbf{s}_j^{[l]\dagger}(t). \quad (28)$$

However, because of finite sample size, its empirical counterpart does not hold; that is, $\overline{\mathbf{S}}_{ij}^{[k,l]} \neq \mathbf{0}_{m_i \times m_j}$. In this section, we develop the error analysis in the regime of small errors; that is, we analyze the error terms $\mathcal{E} = \{\mathcal{E}^{[k]}\}_{k=1}^K$ at first-order in $\overline{\mathbf{S}}_{ij}^{[k,l]}$, when asymptotic conditions (Sec. III-C) hold.

The first-order expansion of the estimating equations (18) yields (see Appendix F) $KN(N-1)$ equations that can be written pairwise, for each $i \neq j$ and all k , as

$$\begin{aligned} -[\mathbf{S}_{ii}^{-1} \overline{\mathbf{S}}_{ij}]_{kk} &= \sum_{l=1}^K [\mathbf{S}_{ii}^{-1}]_{kl} \mathcal{E}_{ij}^{[l]} \mathbf{S}_{jj}^{[l,k]} + (\mathcal{E}_{ij}^{[k]})^\dagger + \Omega\left(\frac{1}{T}\right) \\ -[\mathbf{S}_{jj}^{-1} \overline{\mathbf{S}}_{ji}]_{kk} &= \sum_{l=1}^K [\mathbf{S}_{jj}^{-1}]_{kl} \mathcal{E}_{ji}^{[l]} \mathbf{S}_{ii}^{[l,k]} + (\mathcal{E}_{ji}^{[k]})^\dagger + \Omega\left(\frac{1}{T}\right) \end{aligned} \quad (29)$$

where $[\cdot]_{kl}$ stands for the (k, l) th block of the term in brackets in the appropriate partition. Equation set (29) shows that asymptotically, for each pair of components ($i \neq j$), the error terms $\{\mathcal{E}_{ij}^{[k]}, \mathcal{E}_{ji}^{[k]}\}_{k=1}^K$ are related to the corresponding pair of matrices $(\overline{\mathbf{S}}_{ij}, \overline{\mathbf{S}}_{ji})$ that represents the error in the decorrelation of different groups of dependent sources. Such a pairwise decoupling is customary in the asymptotic analysis of BSS algorithms, e.g., [43, Theorem 11] [17], [35], [44], [45].

Using the $\text{vec}\{\cdot\}$ operator, (29) can be rewritten, for each pair $i \neq j$, as

$$-\mathbf{g} = \mathcal{H}\mathbf{e} + \Omega\left(\frac{1}{T}\right) \quad (30)$$

where \mathbf{e} and \mathbf{g} are $2Km_i m_j \times 1$ vectors,

$$\mathbf{e} = \begin{bmatrix} \mathbf{e}_{ij} \\ \mathbf{e}_{ji} \end{bmatrix}, \quad \mathbf{e}_{ij} = \begin{bmatrix} \text{vec}\{\mathcal{E}_{ij}^{[1]}\} \\ \vdots \\ \text{vec}\{\mathcal{E}_{ij}^{[K]}\} \end{bmatrix}, \quad (31)$$

$$\mathbf{g} = \begin{bmatrix} \mathbf{g}_{ij} \\ \mathbf{g}_{ji} \end{bmatrix}, \quad \mathbf{g}_{ij} = \begin{bmatrix} \text{vec}\{[\mathbf{S}_{ii}^{-1}\bar{\mathbf{S}}_{ij}]_{11}\} \\ \vdots \\ \text{vec}\{[\mathbf{S}_{ii}^{-1}\bar{\mathbf{S}}_{ij}]_{KK}\} \end{bmatrix} \quad (32)$$

and

$$\mathcal{H} = \begin{bmatrix} \mathbf{S}_{jj} \odot \mathbf{S}_{ii}^{-1} & \mathbf{I}_K \otimes \mathcal{T}_{m_j, m_i} \\ \mathbf{I}_K \otimes \mathcal{T}_{m_i, m_j} & \mathbf{S}_{ii} \odot \mathbf{S}_{jj}^{-1} \end{bmatrix} \quad (33)$$

is a $2Km_i m_j \times 2Km_i m_j$ matrix.

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

is a $Km_i m_j \times Km_i m_j$ matrix whose (k, l) th block according to the partition $m_i m_j \mathbf{1}_K$ is $\mathbf{S}_{jj}^{[k,l]} \otimes [\mathbf{S}_{ii}^{-1}]_{kl}$. In the transition from (29) to (30), (31), (32) and (33) we have used the identities (46) in Appendix A. The commutation matrix [46] $\mathcal{T}_{P,Q} \in \mathbb{R}^{PQ \times PQ}$ is such that $\text{vec}\{\mathbf{M}^\dagger\} = \mathcal{T}_{P,Q} \text{vec}\{\mathbf{M}\}$ for any $\mathbf{M} \in \mathbb{R}^{P \times Q}$. More properties of the commutation matrix can be found in Appendix A. Assuming that \mathcal{H} is invertible⁴, we rewrite (30) as

$$\mathbf{e} = -\mathcal{H}^{-1}\mathbf{g} + \Omega\left(\frac{1}{T}\right) \quad i \neq j. \quad (34)$$

Equation (34) shows how the empirical correlation between the sources, that is, the fact that $\bar{\mathbf{S}}_{ij}$ is non-zero in finite sample size, results in non-zero terms \mathcal{E} . Equation (34) is the desired closed-form, first-order expression for the error terms in (27).

D. Closed-Form Expressions for $\text{Cov}(\text{vec}\{\delta\mathbf{P}_i^{[k]}\})$ and MSE

The first step in expressing $\text{Cov}(\text{vec}\{\delta\mathbf{P}_i^{[k]}\})$ as a function of the model parameters is vectorizing (27). Using identity (46c) we obtain

$$\text{vec}\{\delta\mathbf{P}_i^{[k]}\} = (\mathbf{A}^{-[k]\dagger} \otimes \mathbf{A}^{[k]}) \sum_{j \neq i}^N ((\mathbf{E}_j \otimes \mathbf{E}_i) \text{vec}\{\mathcal{E}_{ij}^{[k]}\} - (\mathbf{E}_i \otimes \mathbf{E}_j) \text{vec}\{\mathcal{E}_{ji}^{[k]}\}) + \Omega\left(\frac{1}{T}\right). \quad (35)$$

The covariance of $\text{vec}\{\delta\mathbf{P}_i^{[k]}\}$ can be expressed as

$$\text{Cov}(\text{vec}\{\delta\mathbf{P}_i^{[k]}\}) = (\mathbf{A}^{-[k]\dagger} \otimes \mathbf{A}^{[k]}) \quad (36)$$

$$(\mathbf{M}_{11} + \mathbf{M}_{12} + \mathbf{M}_{21} + \mathbf{M}_{22})(\mathbf{A}^{-[k]} \otimes \mathbf{A}^{[k]\dagger}) + O\left(\frac{1}{T^{3/2}}\right)$$

⁴In the sequel, we assume that \mathcal{H} is invertible. The invertibility of \mathcal{H} is associated with the uniqueness of the model. Further discussion of this point is beyond the scope of this paper.

where

$$\mathbf{M} \triangleq \begin{bmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{21} & \mathbf{M}_{22} \end{bmatrix} \triangleq \sum_{j \neq i}^{N_c} \mathbf{F}_{ij} \text{Cov} \left(\begin{bmatrix} \text{vec}\{\mathcal{E}_{ij}^{[k]}\} \\ \text{vec}\{\mathcal{E}_{ji}^{[k]}\} \end{bmatrix} \right) \mathbf{F}_{ij}^\dagger, \quad (37)$$

$$\mathbf{F}_{ij} \triangleq \begin{bmatrix} \mathbf{E}_j \otimes \mathbf{E}_i & \mathbf{0} \\ \mathbf{0} & -\mathbf{E}_i \otimes \mathbf{E}_j \end{bmatrix} \text{ and}$$

$$\text{Cov}(\mathbf{e}) = \frac{1}{T} \mathcal{H}^{-1} \text{Cov}(\mathbf{g}) \mathcal{H}^{-1} = \frac{1}{T} \mathcal{H}^{-1} + O\left(\frac{1}{T^{3/2}}\right), \quad (38)$$

as we now explain. Eq. (38) follows from Appendix G, where we show that $\text{Cov}(\mathbf{g}) = \frac{1}{T} \mathcal{H}$. Eq. (38) fully characterizes the covariance terms in (37). The asymptotic error term in (38) and (36) and the elimination of the double sum in (37) follow from Appendix G, which implies, combined with (34), that the pairs $(\mathcal{E}_{ij}^{[k]}, \mathcal{E}_{ai}^{[k]})$ and $(\mathcal{E}_{ij}^{[k]}, \mathcal{E}_{ia}^{[k]})$ are asymptotically uncorrelated $\forall a \neq j$. We have thus obtained a closed-form expression (36) for the covariance of all the entries of $\delta\mathbf{P}_i^{[k]}$ that is a function only of the model parameters \mathcal{A} , \mathbf{S} and \mathcal{M} and that is invariant to the arbitrary scaling between \mathcal{A} and \mathbf{S} . This expression can be substituted in (25) for a closed-form expression of the MSE. Further simplification of the MSE can be obtained by using property (46d) and the block-diagonal structure of $\mathbf{S}^{[k,k]}$. This allows us to write

$$\text{MSE}_i^{[k]} = \text{tr}\{(\mathbf{X}^{[k,k]} \otimes \mathbf{I}_M)(\mathbf{A}^{-[k]\dagger} \otimes \mathbf{A}^{[k]}) (\mathbf{M}_{11} + \mathbf{M}_{22})(\mathbf{A}^{-[k]} \otimes \mathbf{A}^{[k]\dagger})\} + O\left(\frac{1}{T^{3/2}}\right) \quad (39)$$

where the terms that depend on \mathbf{M}_{12} and \mathbf{M}_{21} , defined in (37), vanish.

E. FIM, CRLB and MMSE

For samples that follow the Gaussian model in Sec. II, the results in this section have the following interpretation. Eq. (36) is the asymptotically achievable CRLB on the estimation of $\mathbf{P}_i^{[k]}$ and its inverse is the FIM. Accordingly, (39) is the MMSE.

We point out that all the derivations in Sec. IV and the related appendices do not rely on the Gaussian distribution nor on statistics of order larger than two. Therefore, all the equalities in this section hold also for non-Gaussian observations. That is, (36) and (39) still reflect the error covariance and MSE if we apply the methods in Sec. III; however, the CRLB, FIM and MMSE interpretation no longer applies.

V. WELL-POSEDNESS OF THE JISA SOS MODEL

We now discuss conditions under which blind identification of the component subspaces is possible.

1) *Degrees of Freedom*: Let us compare the number of degrees of freedom in the model with the number of constraints in the data, as in [35, Sec. V.A]. The data are represented by a symmetric $L \times L$ matrix, such that the model tries to fit $N_{\text{data}} = \frac{1}{2}L(L-1)$ scalar numbers.

The model consists of K $M \times M$ mixing matrices and N $n_i \times n_i$ source covariance matrices. These provide $N_{\text{model}} = K(M^2 - \sum_{i=1}^N m_i^2) + \frac{1}{2} \sum_{i=1}^N n_i(n_i - 1)$ effective free scalar

parameters, when scale ambiguities are taken into account. It is immediate to verify that

$$N_{\text{data}} - N_{\text{model}} = \frac{1}{2}(K-2)(M^2 - \sum_{i=1}^N m_i^2) \quad (40)$$

Hence, as soon as $K > 1$, there are as many scalar statistics as free parameters in the model. The same calculation shows that imposing statistical independence of all pairs $(\mathbf{s}_i^{[k]}, \mathbf{s}_i^{[k' \neq k]})$ yields a model that is never blindly identifiable using SOS. This result is not surprising, since such a model amounts to K separate BSS problems.

2) *Uniqueness and Identifiability*: The previous argument makes it plausible that for randomly chosen source covariance matrices, the component subspaces can be uniquely identified. In fact, and generalizing IVA [47], the uniqueness of the JISA model can be preserved even if not all entries of $\mathbf{s}_i(t)$ are mutually statistically dependent; further discussion of this point is beyond the scope of this paper. In the following, we assume that the uniqueness conditions are satisfied. It is only for the simplicity of presentation that, in this paper, we assume that all corresponding components are mutually dependent, i.e., no entries fixed to zero in \mathbf{S}_{ii} . Since the mixing matrices are assumed to be invertible, uniqueness of the decomposition implies identifiability of the model.

VI. ALGORITHM

The error analysis in Sec. IV gives rise to a QN algorithm, which is given in pseudocode in Algorithm 1. In [6], a RG algorithm based on the derivation in Sec. III was introduced. Both algorithms maximize the same contrast function (12) and thus obtain the same MSE, up to numerical precision. The QN algorithm is based on approximations of the second-order derivatives and thus enjoys improved convergence properties such as a much smaller number of iterations.

Algorithm 1 An iterative algorithm for JISA

```

1: function JISA( $\bar{\mathbf{X}}, \Phi, \mathbf{A}_{\text{init}}, \mathbf{m}, \text{threshold}, K$ )
2:    $\mathbf{A} \leftarrow \mathbf{A}_{\text{init}}, \mathbf{R} \leftarrow \bar{\mathbf{X}}$  ▷ Init
3:   while  $\|\nabla C\| > \text{threshold}$  do
4:     for  $i=2:N, j=1:i-1$  do ▷ Sweep on  $(i, j \neq i)$ 
5:        $\mathbf{g} \leftarrow \begin{bmatrix} \text{vecbd}_{m_i \mathbf{1}_K \times m_j \mathbf{1}_K} \{\mathbf{R}_{ii}^{-1} \mathbf{R}_{ij}\} \\ \text{vecbd}_{m_j \mathbf{1}_K \times m_i \mathbf{1}_K} \{\mathbf{R}_{jj}^{-1} \mathbf{R}_{ji}\} \end{bmatrix}$  ▷ (32)
6:        $\mathcal{H} \leftarrow \begin{bmatrix} \mathbf{R}_{jj} \odot \mathbf{R}_{ii}^{-1} & \mathbf{I}_K \otimes \mathcal{T}_{m_j, m_i} \\ \mathbf{I}_K \otimes \mathcal{T}_{m_i, m_j} & \mathbf{R}_{ii} \odot \mathbf{R}_{jj}^{-1} \end{bmatrix}$  ▷ (33)
7:       Evaluate  $\mathcal{E}_{ij}^{[k]}, \mathcal{E}_{ji}^{[k]}, k=1, \dots, K$  ▷ (34)
8:       Set  $\{\mathcal{E}_{ij}^{[k]}, \mathcal{E}_{ji}^{[k]}\}_{k=1}^K$  in  $\mathcal{E} = \bigoplus_{k=1}^K \mathcal{E}^{[k]}$ 
9:        $\mathbf{T} \leftarrow \mathbf{I} + \mathcal{E}$  ▷  $\mathcal{E}_{ii}^{[k]} = 0$ 
10:       $\mathbf{R} \leftarrow \mathbf{T}^{-1} \mathbf{R} \mathbf{T}^{-\dagger}$ 
11:       $\mathbf{A} \leftarrow \mathbf{A} \mathbf{T}$  ▷ For output only
12:     end for
13:      $\mathbf{Y} \leftarrow \Phi \mathbf{A}^{-1} \mathbf{R} \mathbf{A}^{-\dagger} \Phi^\dagger$ 
14:      $\nabla C \leftarrow \text{bdiag}_k \{\Phi^\dagger \text{bdiag}_n^{-1} \{\mathbf{Y}\} \mathbf{Y} \Phi\} - \mathbf{I}$  ▷ (17)
15:   end while
16:   return  $\mathbf{A}$ 
17: end function

```

In line 5 of Algorithm 1 we introduced the operator $\text{vecbd}_{\alpha \times \beta} \{\mathbf{X}\} \triangleq [\text{vec}^\dagger \{\mathbf{X}_{11}\} \dots \text{vec}^\dagger \{\mathbf{X}_{KK}\}]^\dagger$, where $\text{vecbd}_{\alpha \times \beta} \{\mathbf{X}\}$ is a vector that consists only of the (vectorized) entries of the block-diagonal blocks of matrix \mathbf{X} . Matrices \mathbf{X}_{kk} are the blocks on the main diagonal of \mathbf{X} where the rows of \mathbf{X} are partitioned according to α and the columns by β .

Naturally, the elements of the gradient \mathbf{g} (line 5 of Algorithm 1) correspond to certain entries of $\nabla C(\mathbf{A})$ (line 14 of Algorithm 1). However, since \mathbf{g} changes within the loop, then $\nabla C(\mathbf{A})$, which consists of gradient values for *all* pairs (i, j) , has to be fully re-calculated again at the exit. In line 8, it follows from (18) that $\text{bdiag}_k \{\nabla C(\mathbf{A})\}$ is invariant to changes in $\bar{\mathbf{X}}$; therefore, we set $\mathcal{E}_{ii}^{[k]} = 0$.

VII. NUMERICAL VALIDATION

In this section, we validate the theoretical results that were derived in previous sections. The main contribution of this paper is illustrated in Fig. 3–4, in which we validate, using the QN Algorithm 1, that the closed-form expression for the MSE (25), developed in Sec. IV, indeed predicts correctly the empirical MSE (21). Fig. 3–4 also illustrate the gain from using the true multidimensional approach (JISA) over analysis in one-dimensional terms (IVA) followed by clustering the output into the appropriate subspaces. Finally, Fig. 4 validates that indeed our SOS analysis holds equally well for non-Gaussian data, as explained in Sec. IV-E. The proper functioning of the QN algorithm, in terms of convergence and small number of iterations, is demonstrated in sec. VII-A. Its component-separation capacities, as well as the theoretical validation of the MSE prediction, are illustrated in Sec. VII-B.

In all the following numerical experiments, the real positive definite matrices \mathbf{S}_{ii} are drawn as $\mathbf{S}_{ii} = \text{diag}^{-\frac{1}{2}} \{\mathbf{U} \mathbf{A} \mathbf{U}^\dagger\} \mathbf{U} \mathbf{A} \mathbf{U}^\dagger \text{diag}^{-\frac{1}{2}} \{\mathbf{U} \mathbf{A} \mathbf{U}^\dagger\}$, where $\mathbf{U} \mathbf{A} \mathbf{V}^\dagger$ is the singular value decomposition (SVD) of a $K m_i \times K m_i$ matrix whose independent and identically distributed (i.i.d.) entries $\sim \mathcal{N}(0, 1)$. The corresponding samples are created by right-multiplying the transpose of the Cholesky factorization of $\bar{\mathbf{S}}_{ii}$ with $K m_i \times T$ i.i.d. zero-mean, unit-variance numbers: Gaussian in Sec. VII-A and Fig. 3, Gaussian mixture (GM) with peaks centred at ± 4 in Fig. 4. The stopping threshold is set to 10^{-6} , and $T = 10^4$. For “backwards compatibility” we verified that our algorithm achieves the same optimum as the JBSS SOS algorithm in [41] for $\mathbf{m} = \mathbf{1}_{M \times 1}$, up to acceptable numerical variations.

Since the component estimates are invariant to block-diagonal scaling (Sec. II), we are concerned only about permutation issues. Permutation that is not properly corrected is interpreted as significant errors in component reconstruction. As follows from Sec. II, JISA usually does not suffer from permutation errors, at least as long as deviations from the model due to various errors are sufficiently small. Therefore, our major concern is the global arbitrary permutation, that implies that even if the algorithm has converged properly and found the global minimum of the contrast function, the evaluation of the MSE per component will fail if this arbitrary permutation is not corrected w.r.t. the original ordering of the

components. Since finding an arbitrary permutation requires enumerating on all $\frac{M!}{\prod_{i=1}^N (m_i!)}$ options and thus impractical, we bypassed this issue via the safety margins that we took in choosing \mathbf{A}_{init} w.r.t. its true value. More specifically, we set $\mathbf{A}^{[k]} = p\mathbf{\Upsilon} + (1-p)\mathbf{I}$ and $\mathbf{A}_{\text{init}}^{[k]} = p\mathbf{\Upsilon} + (1-p)\mathbf{A}$, $p = 0.2$, unless stated otherwise. The entries of $\mathbf{\Upsilon}$ are $\sim \mathcal{N}(0, 1)$ i.i.d. and drawn anew for each $\mathbf{A}^{[k]}$ and $\mathbf{A}_{\text{init}}^{[k]}$.

A. Convergence

We now illustrate the convergence properties of the algorithm. In this example, $\mathbf{m} = [1, 2, 3]^\dagger$, $K = 4$. Only \mathbf{A}_{init} varies at each of the 40 Monte Carlo (MC) trials, \mathbf{A} , \mathbf{S} and $\widetilde{\mathbf{S}}$ (the empirical counterpart of \mathbf{S}) remain fixed. Fig. 2 shows the value of the contrast function (up) and number of iterations (bottom) when the stopping criterion is achieved. At each trial, the algorithm is run twice on the same data, in two modes: in the first mode, the input parameter \mathbf{m} (Algorithm 1 line 1) reflects the true model structure. In the second mode, the input parameter is set to $\mathbf{1}_{M \times 1}$, a vector of all ones. The latter implies that the algorithm is ignorant of the true multidimensional structure of the data and instead tries to fit it to a one-dimensional model. This corresponds to *applying*, in a first step, a classical SOS IVA/JBSS model to the data, with $M = 6 = N$ independent sources at each mixture. In a second step, we cluster the output into the correct N multidimensional components before evaluating the contrast function. We denote this two-step approach “mismodeling” [17].

It is clear from Fig. 2 (up) that the QN algorithm indeed converges to the same optimum (up to small variations due to finite threshold) regardless of init, as expected. It is interesting to note that in adverse conditions, that is, mismodeling, the variance is slightly larger. We have noted that this property becomes more pronounced as the mismodeling further departs from the true data model. In addition, we note in Fig. 2 (bottom) that the mismodeled case requires a larger number of iterations to converge. This, too, is expected, since the algorithm is trying to block-diagonalize $\widetilde{\mathbf{S}}$ into smaller blocks than is actually possible (by definition of the multidimensional model) and thus doing unnecessary work. Comparison (not presented in this paper) with the RG JISA algorithm [6] has validated that the QN indeed converges in a much smaller number of iterations, as expected.

B. Component Separation

We now evaluate the component separation quality of the proposed model and algorithm. In the following experiment, we run multiple trials for fixed \mathbf{S} , \mathbf{A} , and \mathbf{A}_{init} . Only $\widetilde{\mathbf{S}}$ varies. In Fig. 3, $\mathbf{m} = [6, 5, 1]^\dagger$, $K = 5$ and Gaussian samples. In Fig. 4, $\mathbf{m} = [3, 5, 4]^\dagger$, $K = 6$ and the samples are generated from numbers with a GM distribution, as explained at the beginning of Sec. VII. Note that right-multiplying non-Gaussian numbers with a Cholesky factorization of $\widetilde{\mathbf{S}}_{ii}$ changes their distribution; however, it is still non-Gaussian.

As in Sec. VII-A, we compare JISA with JBSS. In the latter case, we cluster the output into the correct N multidimensional components before evaluating the MSE. In JBSS-mismodeling, we have noted that as the true \mathbf{m} departs

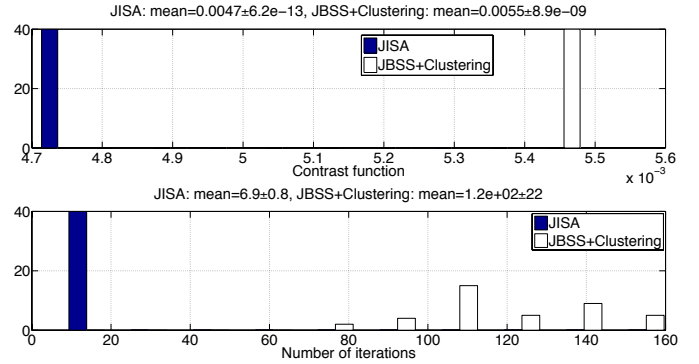


Fig. 2. Convergence of the QN algorithm. Histogram of the value of the contrast function (up) and number of iterations (bottom) when the stopping criterion is achieved.

from $\mathbf{1}_M$, at finite sample size, the algorithm becomes more sensitive to the init, in the sense that the global permutation matrix deviates from that induced by the original \mathbf{A}_{init} . In order to overcome this *for the error-analysis validation purpose*, we chose a more strict initialization strategy in which in the first attempt \mathbf{A}_{init} is taken from the output of the JISA run, and if the empirical MSE indicates global permutation that does not fit the original components (recall explanation at the introduction to Sec. VII), new \mathbf{A}_{init} are generated according to the original procedure until no permutation issues are detected. For each trial we evaluate the normalized empirical MSE (21). Fig. 3–4 illustrate our results. Subplot i corresponds to component i . Each subplot shows the results of two different runs of the algorithm on the same data: once as true JISA with the correct data model, and once mismodeled, as explained earlier in this section. For each of the two runs of the algorithm, the mean and standard deviation (std) of $\widehat{\text{MSE}}_i$ are written above the corresponding subplot, together with the predicted JISA MSE (25). The histograms represent 200 MC repetitions of this experiment. The three MSE values are marked on the histograms.

C. Discussion

Fig. 3–4 validate that the closed-form MSE (25) indeed predicts the empirical value, both for Gaussian and non-Gaussian data, as expected. This also serves as an implicit validation that we are indeed in the small-errors regime. The small values of the normalized $\widehat{\text{MSE}}_i$ in Fig. 3–4 confirm that the components have been properly separated. Fig. 3–4 show that for all three components, there is a significant improvement (decrease) in MSE when the correct model is used w.r.t. the mismodeled scenario, both in the mean and in the std. This observation conforms with previous results on multidimensional components [17], [35]. The larger std for the mismodeling scenario, along with the larger std and larger number of required iterations, observed in Fig. 2, are all due to the considerable model misfit.

On the other hand, a possible advantage of the JBSS+clustering approach, that still needs to be verified from further numerical examinations, is that it is more robust than JISA to initialization in the sense of convergence to the global

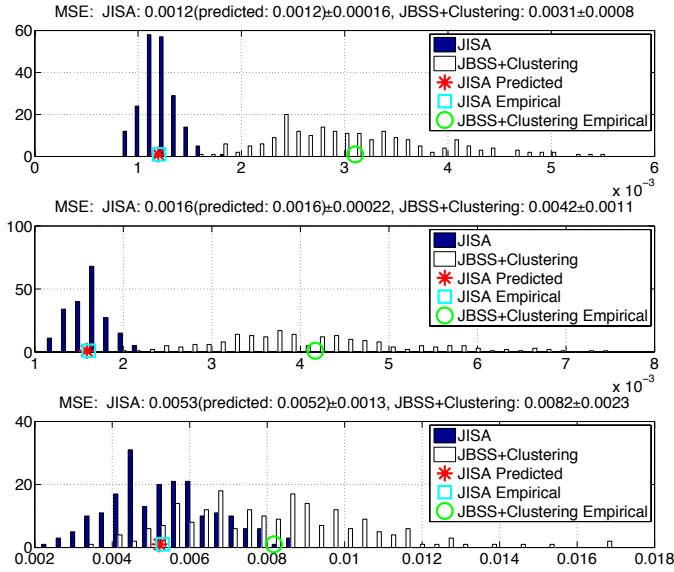


Fig. 3. Component separation. Histogram of the normalized empirical MSE. Subplots correspond to components with dimensions 6, 5 and 1, respectively. $K = 5$, Gaussian data.

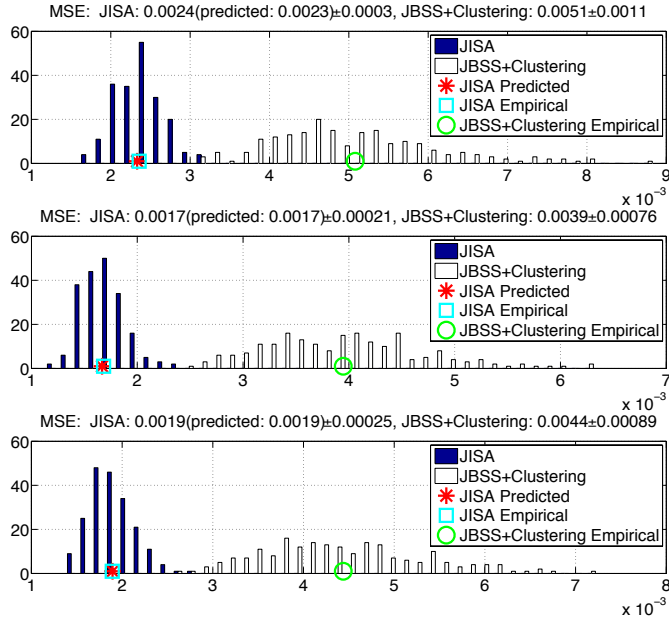


Fig. 4. Component separation. Histogram of the normalized empirical MSE. Subplots correspond to components with dimensions 3, 5 and 4, respectively. $K = 6$, non-Gaussian data.

minimum of the contrast function. We conjecture that since JISA works in blocks, if the permutation induced by the initialization does not perfectly fit the arbitrary permutation of the data, then the algorithm will mix elements of different blocks and thus will not converge to the global optimum. The latter problem is prevalent also in other block decompositions such as joint block diagonalization (JBD) [15].

It should be emphasized that the focus of this paper is on *validating* the theoretical error analysis. Therefore, numerical and practical issues such as: identifying the global permutation, number and dimension of the latent sources, in the

absence of additional information, local minima, and choice of proper initialization, are beyond the scope of this work.

VIII. DISCUSSION

A. Possible Extensions to the JISA Model

As explained in footnote 1, the basic JISA formulation in Sec. II is rather basic. It assumes that the matrix and source subspace partitions all have the same dimensions in all mixtures, and a determined (invertible) mixing. Obviously, one can define a more general JISA model with different dimensions in each mixture, as explicitly stated in [34]. We deliberately chose this simplification because it is sufficiently simple to keep notations and derivations clear and easily tractable yet at the same time sufficiently rich to encompass the essential properties of JISA. It is easy to verify that our results and algorithm generalize straightforwardly to multidimensional components of different dimensions, as long as $\mathbf{A}^{[k]}$, now of different dimensions, remain invertible. That is, $m_i^{[k]}$ instead of m_i , $M^{[k]}$ instead of M , and so on. There is no need for new derivations, only putting the right indices in the right places. We avoid this in our paper since it clutters the notations and does not add any fundamental new insights. A generalization of IVA to matrices of different dimensions has been mentioned in [28].

As mentioned in footnote 1 and Sec. V, one can relax the statistical dependence across mixtures, as long as uniqueness is preserved. This means that one can have also $N^{[k]}$ instead of N ; this can be accommodated in our model by setting certain components to zero, to imply non-corresponding elements. Uniqueness can still hold, in certain conditions, as one can easily verify numerically. This result can be proved (though beyond the scope of this paper) and is a generalization of this property for IVA [30].

Finally, it is straightforward to fuse the JISA model, which is based on the linked-between-datasets diversity, with the nonstationarity diversity [17], [35]; this will generalize JISA to non-stationary or coloured sources, in a straightforward way as suggested by [32] and [28] for the one-dimensional case of IVA. These are trivial extensions that do not require any fundamental change of our derivations or algorithm, only a weighted average of the likelihood.

B. Alternative Algebraic Representations of JISA

It is interesting to note that JISA has two alternative algebraic representations. This is analogous to the multidimensional nonstationary SOS-based BSS [17], [35], which can be treated either as JBD or as a block term decomposition (BTD) [18].

For the first algebraic form, generalizing IVA/JBSS [28], JISA can be regarded as a coupled cyclic JBD model, as follows. Let

$$\mathbf{S}^{[k,l]} \triangleq E\{\mathbf{s}^{[k]}(t)\mathbf{s}^{[l]\dagger}(t)\} = \begin{bmatrix} \mathbf{S}_{11}^{[k,l]} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{S}_{NN}^{[k,l]} \end{bmatrix} \quad (41)$$

where $k, l \in \{1, \dots, K\}$ and

$$\mathbf{S}_{ij}^{[k,l]} = E\{\mathbf{s}_i^{[k]}(t)\mathbf{s}_j^{[l]\dagger}(t)\} \in \mathbb{R}^{m_i \times m_j}. \quad (42)$$

The linear model (1) implies that

$$\mathbf{X}^{[k,l]} = \mathbf{A}^{[k]}\mathbf{S}^{[k,l]}\mathbf{A}^{[l]\dagger} \quad \forall k, l \quad (43)$$

where

$$\mathbf{X}^{[k,l]} = E\{\mathbf{x}^{[k]}(t)\mathbf{x}^{[l]\dagger}(t)\}. \quad (44)$$

Eq. (43) can be rewritten as

$$\mathbf{S}^{[k,l]} = \mathbf{A}^{-[k]}\mathbf{X}^{[k,l]}\mathbf{A}^{-[l]\dagger} \quad \forall k, l \quad (45)$$

Eq. (45) alludes to the desired new coupled JBD problem, which can be regarded as a ‘‘cyclic’’ form, since the K matrices on the right-hand side (RHS) and left-hand side (LHS) come from the same pool of K matrices. Note the analogy to JBD, in which the RHS and LHS are either identical (symmetric) or different (asymmetric) [18].

The second algebraic form of JISA is the one in Sec. II, which can be rephrased as symmetric block-diagonalization of a matrix by a permuted block-diagonal matrix.

C. Link With JBD and Nonstationary Multidimensional BSS

As briefly mentioned in Sec. I, in this paper, we have used the same derivation guidelines as in [17], [35]. This is not surprising, since the two models can be regarded as two different variations on a more general multi-set multidimensional BSS scenario. In JISA, the added diversity, or link between K disparate mixtures (1), is an additional assumption of joint pdf between corresponding sources across mixtures, i.e., $\mathbf{S}_{ij}^{[k]}$ generally non-zero. In multidimensional nonstationary BSS, the diversity, or the property that generates the link, is forcing $\mathbf{A}^{[k]} = \mathbf{A}$ (which reduces the number of degrees of freedom) and then compensating for the lack of equations in the SOS case (or for Gaussian data) with a nonstationarity assumption on the sources. The structural similarity between the contrast function (12) and Hessian in (33) and their counterparts in [17], [35] expresses the analogy between these two models.

APPENDIX A SOME ALGEBRAIC PROPERTIES

For ease of reference, we list some useful algebraic properties. Properties that are not proved below can be found in [46], [48], [49].

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

$$(\mathbf{N} \otimes \mathbf{M})(\mathbf{P} \otimes \mathbf{Q}) = \mathbf{NP} \otimes \mathbf{MQ} \quad (46a)$$

$$(\mathbf{N} \otimes \mathbf{M})^\dagger = \mathbf{N}^\dagger \otimes \mathbf{M}^\dagger \quad (46b)$$

$$\text{vec}\{\mathbf{MQN}\} = (\mathbf{N}^\dagger \otimes \mathbf{M})\text{vec}\{\mathbf{Q}\} \quad (46c)$$

$$\text{tr}\{\mathbf{PQ}\} = \text{tr}\{\mathbf{QP}\} \quad (46d)$$

$$\text{tr}\{\mathbf{P}^\dagger\mathbf{Q}\} = \text{vec}^\dagger\{\mathbf{P}\}\text{vec}\{\mathbf{Q}\} \quad (46e)$$

$$\det(\mathbf{MN}) = \det(\mathbf{NM}). \quad (46f)$$

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

$$\mathcal{T}_{M,N}(\mathbf{N} \otimes \mathbf{M}) = (\mathbf{M} \otimes \mathbf{N})\mathcal{T}_{P,Q}. \quad (47a)$$

Property A.1. Let $\mathbf{L}, \mathbf{M}, \mathbf{N}$ be square matrices. Then $\text{tr}\{\mathbf{MLN}^\dagger\} = \text{tr}\{(\mathbf{L}^\dagger \otimes \mathbf{I})\text{vec}\{\mathbf{M}\}\text{vec}^\dagger\{\mathbf{N}\}\}$. The proof follows the same steps as [35, Property A.1].

Property A.2. The first-order Taylor expansion of $(\mathbf{M} + \mathbf{\Delta})^{-1}$ about \mathbf{M} , where \mathbf{M} and $\mathbf{\Delta}$ are square matrices, \mathbf{M} and $\mathbf{M} + \mathbf{\Delta}$ invertible, is $(\mathbf{M} + \mathbf{\Delta})^{-1} = \mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{\Delta}\mathbf{M}^{-1} + O(\|\mathbf{\Delta}\|^2)$.

APPENDIX B

ASYMPTOTIC EXPRESSION FOR THE MSE: PROOF OF (25)

Without vectorization, (24) can be equally rewritten as

$$\begin{aligned} \widehat{\text{MSE}}_i^{[k]} &= \text{tr}\{\widehat{\mathbf{X}}^{[k,k]}(\delta\mathbf{P}_i^{[k]})^\dagger\delta\mathbf{P}_i^{[k]}\} \\ &= \text{tr}\{\mathbf{X}^{[k,k]}(\delta\mathbf{P}_i^{[k]})^\dagger\delta\mathbf{P}_i^{[k]}\} \\ &\quad + \text{tr}\{\delta\mathbf{X}^{[k,k]}(\delta\mathbf{P}_i^{[k]})^\dagger\delta\mathbf{P}_i^{[k]}\} \end{aligned} \quad (48)$$

where in the last step we have defined

$$\widehat{\mathbf{X}}^{[k,k]} = \mathbf{X}^{[k,k]} + \delta\mathbf{X}^{[k,k]}. \quad (49)$$

Taking expectation of (48) we obtain that

$$\begin{aligned} E\{\widehat{\text{MSE}}_i^{[k]}\} &= \text{tr}\{\mathbf{X}^{[k,k]}E\{(\delta\mathbf{P}_i^{[k]})^\dagger\delta\mathbf{P}_i^{[k]}\}\} \\ &\quad + E\{\text{tr}\{\delta\mathbf{X}^{[k,k]}(\delta\mathbf{P}_i^{[k]})^\dagger\delta\mathbf{P}_i^{[k]}\}\} \end{aligned} \quad (50)$$

In Appendix C we show that both $\delta\mathbf{X}^{[k,k]}$ and $\delta\mathbf{P}_i^{[k]}$ are $O(\frac{1}{\sqrt{T}})$. Therefore, the second summand on the RHS of (50) is (at worst) $O(\frac{1}{T^{3/2}})$, which concludes our proof.

APPENDIX C

ASYMPTOTIC PROPERTIES OF $\delta\mathbf{X}^{[k,k]}$, $\delta\mathbf{P}_i^{[k]}$, $\delta\mathbf{S}$ AND \mathcal{E}

Under asymptotic conditions ($T \rightarrow \infty$), the sample covariances $\widehat{\mathbf{X}}$ (defined in Sec. III) and $\widehat{\mathbf{S}}$ and the ML estimators $\widehat{\mathbf{A}}^{[k]}$ and $\widehat{\mathbf{P}}_i^{[k]}$ (defined in Sec. III-B) converge, respectively, to \mathbf{X} , \mathbf{S} , $\mathbf{A}^{[k]}$ and $\mathbf{P}_i^{[k]}$, at least in probability. As for the rate of convergence, the entries of $\delta\mathbf{X}$, $\delta\mathbf{S}$, $\mathcal{E}^{[k]}$ (26) and $\delta\mathbf{P}_i^{[k]}$ (22) are zero mean random variables with a standard deviation proportional to $1/\sqrt{T}$. For $\delta\mathbf{X}$ and $\delta\mathbf{S}$ this follows from the central limit theorem. For the ML estimation errors $\mathcal{E}^{[k]}$ and $\delta\mathbf{P}_i^{[k]}$ this is due to sample independence. The fact that the entries of $\mathcal{E}^{[k]}$ decrease (asymptotically) with T at the same rate as the entries of $\delta\mathbf{P}_i^{[k]}$ can also be deduced from (27). Assuming that asymptotic conditions hold, then $\widehat{\mathbf{P}}_i^{[k]} \cong \mathbf{P}_i^{[k]}$.

APPENDIX D

INVARIANCE OF THE ESTIMATING EQUATIONS TO SCALE AMBIGUITY OR EQUIVALENCE CLASS OF THE SOLUTIONS TO THE ESTIMATING EQUATIONS (18)

Given the existence of a set \mathcal{A} that satisfies the estimating equations (18) and thus also achieves the minimum of the contrast function (12), we now discuss its equivalence class. That is, the subspace of matrices its generates that also satisfy (18) and minimize (12).

Using the fact that the $\text{bdiag}_b\{\cdot\}$ operator commutes with any $\mathbf{\Lambda}, \mathbf{\Lambda}' \in \mathcal{B}_b$ in the following manner,

$$\text{bdiag}_b\{\mathbf{\Lambda}\mathbf{M}\mathbf{\Lambda}'\} = \mathbf{\Lambda}\text{bdiag}_b\{\mathbf{M}\}\mathbf{\Lambda}', \quad \forall \mathbf{M} \in \mathbb{R}^{M \times M}, \quad (51)$$

and given

$$\Phi \left(\bigoplus_{k=1}^K \left(\bigoplus_{i=1}^N \mathbf{Z}_{ii}^{[k]} \right) \right) = \left(\bigoplus_{i=1}^N \left(\bigoplus_{k=1}^K \mathbf{Z}_{ii}^{[k]} \right) \right) \Phi, \quad (52)$$

(9), (12) (46d) and (46f), we obtain that $C(\mathbf{A}) = C(\mathbf{A}\mathbf{\Lambda}'')$, where $\mathbf{\Lambda}'' \in \mathcal{B}_{1_K \otimes \mathbf{m}}$ coincides with the scale ambiguity of the model (Section II). Similar steps show that the estimating equations (18) are invariant to $\mathbf{A} \rightarrow \mathbf{A}\mathbf{\Lambda}''$.

APPENDIX E DERIVATION OF (27)

In this section, we obtain (27). The proof follows steps similar to those in [17, Appendix B]. Substituting (26) in (6),

$$\begin{aligned} \widehat{\mathbf{P}}_i^{[k]} &= \mathbf{A}^{[k]}(\mathbf{I} + \mathcal{E}^{[k]})^{-1} \mathbf{\Lambda}^{[k]} \mathbf{E}_i \mathbf{E}_i^\dagger \mathbf{\Lambda}^{-[k]} (\mathbf{I} + \mathcal{E}^{[k]}) \mathbf{A}^{-[k]} \\ &= \mathbf{A}^{[k]}(\mathbf{I} + \mathcal{E}^{[k]})^{-1} \mathbf{E}_i \mathbf{E}_i^\dagger (\mathbf{I} + \mathcal{E}^{[k]}) \mathbf{A}^{-[k]} \\ &= \mathbf{P}_i^{[k]} + \mathbf{A}^{[k]} (\mathbf{E}_i \mathbf{E}_i^\dagger \mathcal{E}^{[k]} - \mathcal{E}^{[k]} \mathbf{E}_i \mathbf{E}_i^\dagger) \mathbf{A}^{-[k]} \\ &\quad + O(\|\mathcal{E}^{[k]}\|^2). \end{aligned} \quad (53)$$

The first equality in (53) follows from $\mathbf{A}_i^{[k]} = \mathbf{A}^{[k]} \mathbf{E}_i$ and $\mathbf{B}_i^{[k]} = \mathbf{E}_i^\dagger \mathbf{B}^{[k]}$. The second equality is due to the fact that $\mathbf{\Lambda}^{[k]} \in \mathcal{B}_{\mathbf{m}}$. The third equality uses Property A.2 in Appendix A followed by $\mathbf{P}_i^{[k]} = \mathbf{A}^{[k]} \mathbf{E}_i \mathbf{E}_i^\dagger \mathbf{A}^{-[k]}$, which is due to (6). The last transition, from (53) to (27), follows from $\mathcal{E}^{[k]} \mathbf{E}_i = \sum_{j=1}^N \mathbf{E}_j \mathcal{E}_{ji}^{[k]}$ and $\mathbf{E}_i^\dagger \mathcal{E}^{[k]} = \sum_{j=1}^N \mathcal{E}_{ij}^{[k]} \mathbf{E}_j^\dagger$.

APPENDIX F FIRST-ORDER EXPANSION OF THE ESTIMATING EQUATIONS

In this appendix, we show how a first-order expansion of the estimating equations (18) leads to the linear relation (29) between the error terms \mathcal{E} and the sample covariance matrix $\bar{\mathbf{S}}$. We begin by rewriting (18) with $\widehat{\mathbf{A}}$ instead of \mathbf{A} , emphasizing the fact that solutions of (18) are estimates of \mathbf{A} ,

$$\text{bdiag}_{\mathbf{k}} \{ \Phi^\dagger \text{bdiag}_{\mathbf{n}}^{-1} \{ \Phi \widehat{\mathbf{A}}^{-1} \bar{\mathbf{X}} \widehat{\mathbf{A}}^{-\dagger} \Phi^\dagger \} \Phi \widehat{\mathbf{A}}^{-1} \bar{\mathbf{X}} \widehat{\mathbf{A}}^{-\dagger} \} = \mathbf{I}. \quad (54)$$

The link between \mathcal{E} and $\widehat{\mathbf{A}}$ is given in (26). It follows from Appendix D that the estimating equations (54) are invariant to right-multiplication of $\widehat{\mathbf{A}}$ by any matrix in $\mathcal{B}_{1_K \otimes \mathbf{m}}$. Therefore, in the sequel, we omit the scale ambiguity term of (26). It follows from (4), (26) and the above arguments that (54) can be rewritten as

$$\text{bdiag}_{\mathbf{k}} \{ \Phi^\dagger \text{bdiag}_{\mathbf{n}}^{-1} \{ \Phi (\mathbf{I} + \mathcal{E}) \bar{\mathbf{S}} (\mathbf{I} + \mathcal{E})^\dagger \Phi^\dagger \} \Phi (\mathbf{I} + \mathcal{E}) \bar{\mathbf{S}} (\mathbf{I} + \mathcal{E})^\dagger \} = \mathbf{I}. \quad (55)$$

Given the factorization $\bar{\mathbf{S}} = \mathbf{S} + \delta\mathbf{S}$ one has

$$\begin{aligned} (\mathbf{I} + \mathcal{E}) \bar{\mathbf{S}} (\mathbf{I} + \mathcal{E})^\dagger &= \bar{\mathbf{S}} + \mathcal{E}\mathbf{S} + \mathbf{S}\mathcal{E}^\dagger + \Omega\left(\frac{1}{T}\right) \\ &= \mathbf{S} + \Omega\left(\frac{1}{\sqrt{T}}\right), \end{aligned} \quad (56)$$

which is due to the fact that both \mathcal{E} and $\delta\mathbf{S}$ are $\Omega\left(\frac{1}{\sqrt{T}}\right)$, as explained in Appendix C. Left- and right-multiplying (56) by Φ and Φ^\dagger , respectively, applying $\text{bdiag}_{\mathbf{n}}^{-1}\{\cdot\}$ and then Property A.2 in Appendix A, one obtains

$$\text{bdiag}_{\mathbf{n}}^{-1} \{ \Phi (\mathbf{I} + \mathcal{E}) \bar{\mathbf{S}} (\mathbf{I} + \mathcal{E})^\dagger \Phi^\dagger \} = \widetilde{\mathbf{S}}^{-1} + \widetilde{\Theta} \quad (57)$$

where the entries of $\widetilde{\Theta} \in \mathcal{B}_{\mathbf{n}}$ are $\Omega\left(\frac{1}{\sqrt{T}}\right)$. The term within $\text{bdiag}_{\mathbf{k}}\{\cdot\}$ in (55) can now be rewritten as

$$\begin{aligned} &(\mathbf{S}^{-1} + \Theta)(\mathbf{I} + \mathcal{E}) \bar{\mathbf{S}} (\mathbf{I} + \mathcal{E})^\dagger \\ &= \mathbf{S}^{-1} \bar{\mathbf{S}} + \mathbf{S}^{-1} \mathcal{E}\mathbf{S} + \mathcal{E}^\dagger + \Theta\mathbf{S} + \Omega\left(\frac{1}{T}\right) \end{aligned} \quad (58)$$

where $\Theta = \Phi^\dagger \widetilde{\Theta} \Phi$ and

$$\Phi^\dagger (\widetilde{\mathbf{S}}^{-1} + \widetilde{\Theta}) \Phi = \mathbf{S}^{-1} + \Theta. \quad (59)$$

Using (58), the estimating equations can now be rewritten as

$$\begin{aligned} &\text{bdiag}_{\mathbf{k}} \{ \mathbf{S}^{-1} \bar{\mathbf{S}} \} + \text{bdiag}_{\mathbf{k}} \{ \mathbf{S}^{-1} \mathcal{E}\mathbf{S} \} + \mathcal{E}^\dagger \\ &\quad + \text{bdiag}_{\mathbf{k}} \{ \Theta\mathbf{S} \} + \text{bdiag}_{\mathbf{k}} \{ \Omega\left(\frac{1}{T}\right) \} = \mathbf{I} \end{aligned} \quad (60)$$

since $\mathcal{E} \in \mathcal{B}_{\mathbf{k}}$.

It is clear that entries outside $\text{bdiag}_{\mathbf{k}}\{\cdot\}$ do not yield any constraints on \mathcal{E} . It can be further verified that entries on the main diagonal of (60) with block-pattern $\mathbf{1}_K \otimes \mathbf{m}$ are identical on both sides of (60) and thus do not have any effect. It follows that the only non-trivial terms in (54) can be written as a set of $KN(N-1)$ equations

$$\begin{aligned} &[[\mathbf{S}^{-1} \bar{\mathbf{S}}]_{kk}]_{ij} + [[\mathbf{S}^{-1} \mathcal{E}\mathbf{S}]_{kk}]_{ij} + (\mathcal{E}_{ji}^{[k]})^\dagger \\ &\quad + [[\Theta\mathbf{S}]_{kk}]_{ij} + \Omega\left(\frac{1}{T}\right) = \mathbf{0}_{m_i \times m_j}, \quad i \neq j \end{aligned} \quad (61)$$

where blocks indexed by k follow block-pattern \mathbf{k} , those indexed by i and j follow \mathbf{m} , and $[[\mathcal{E}^\dagger]_{kk}]_{ij} = (\mathcal{E}_{ji}^{[k]})^\dagger$.

The next step is to simplify each of the summands in (61). For the first summand,

$$\begin{aligned} [[\mathbf{S}^{-1} \bar{\mathbf{S}}]_{kk}]_{ij} &= \mathbf{E}_i^\dagger \mathbf{J}_k^\dagger \Phi^\dagger \widetilde{\mathbf{S}}^{-1} \widetilde{\Phi} \mathbf{J}_k \mathbf{E}_j \\ &= \mathbf{U}_k^\dagger \mathbf{Y}_i^\dagger \Phi \Phi^\dagger \widetilde{\mathbf{S}}^{-1} \widetilde{\Phi} \Phi^\dagger \mathbf{Y}_j \mathbf{U}_k \\ &= \mathbf{U}_k^\dagger \mathbf{S}_{ii}^{-1} \bar{\mathbf{S}}_{ij} \mathbf{U}_k = [\mathbf{S}_{ii}^{-1} \bar{\mathbf{S}}_{ij}]_{kk} \end{aligned} \quad (62)$$

where the first step uses (7) such that $\mathbf{S}^{-1} \bar{\mathbf{S}} = \Phi^\dagger \widetilde{\mathbf{S}}^{-1} \Phi \Phi^\dagger \widetilde{\mathbf{S}} \Phi = \Phi^\dagger \widetilde{\mathbf{S}}^{-1} \widetilde{\Phi} \Phi$. In the second step of (62) we employ the observation that

$$\mathbf{J}_k \mathbf{E}_j = \Phi^\dagger \mathbf{Y}_j \mathbf{U}_k \quad (63)$$

where \mathbf{U}_k is $n_i \times m_i$ and \mathbf{Y}_i is $L \times n_i$ such that $\mathbf{I}_{n_i} = [\mathbf{U}_1 | \dots | \mathbf{U}_K]$ and $\mathbf{I}_L = [\mathbf{Y}_1 | \dots | \mathbf{Y}_N]$, respectively⁵. The third step uses $[\widetilde{\mathbf{S}}^{-1} \bar{\mathbf{S}}^{-1}]_{ij} = \mathbf{S}_{ii}^{-1} \bar{\mathbf{S}}_{ij}$, which follows from $\widetilde{\mathbf{S}} \in \mathcal{B}_{\mathbf{n}}$. For the second summand in (61), we begin by writing explicitly the term within $[\cdot]_{ij}$,

$$[\mathbf{S}^{-1} \mathcal{E}\mathbf{S}]_{kk} = \sum_{l=1}^K \mathbf{J}_k^\dagger \mathbf{S}^{-1} \mathbf{J}_l \mathcal{E}^{[l]} \mathbf{J}_l^\dagger \mathbf{S} \mathbf{J}_k = \sum_{l=1}^K [\mathbf{S}^{-1}]_{kl} \mathcal{E}^{[l]} \mathbf{S}^{[l,k]}$$

where $\mathcal{E} \triangleq \sum_{l=1}^K \mathbf{J}_l \mathcal{E}^{[l]} \mathbf{J}_l^\dagger$. It follows from (7) that \mathbf{S}^{-1} has the same zero-pattern as \mathbf{S} such that $[\mathbf{S}^{-1}]_{kl} \in \mathcal{B}_{\mathbf{m}}$ and $[[\mathbf{S}^{-1}]_{kl}]_{ii} = [\mathbf{S}_{ii}^{-1}]_{kl}$. Hence,

$$[[\mathbf{S}^{-1}]_{kl} \mathcal{E}^{[l]} \mathbf{S}^{[l,k]}]_{ij} = [[\mathbf{S}^{-1}]_{kl}]_{ii} \mathcal{E}_{ij}^{[l]} \mathbf{S}_{jj}^{[l,k]} = [\mathbf{S}_{ii}^{-1}]_{kl} \mathcal{E}_{ij}^{[l]} \mathbf{S}_{jj}^{[l,k]}$$

which is the desired form. We conclude the derivation of the first equation in (29) by showing that $[[\Theta\mathbf{S}]_{kk}]_{ij} = \mathbf{0}_{m_i \times m_j}$.

⁵Note that \mathbf{U}_k implicitly depends on i ; this is due to the special structure of Φ . The symmetry between the two sides of (63) becomes more obvious when one relaxes m_i to $m_i^{[k]}$, since then \mathbf{E}_j becomes dependent on k .

This follows from noting that $\Theta\mathbf{S} = \Phi^\dagger\tilde{\Theta}\Phi\Phi^\dagger\tilde{\mathbf{S}}\Phi = \Phi^\dagger\tilde{\Theta}\tilde{\mathbf{S}}\Phi$ has the same zero-pattern as \mathbf{S} and by definition, $[[\mathbf{S}]_{kk}]_{ij} = \mathbf{0}_{m_i \times m_j}$. The second equation in (29) is obtained by exchanging i and j .

APPENDIX G CLOSED-FORM EXPRESSION FOR $\text{Cov}(\mathbf{g})$

In this Appendix we derive a closed-form expression for the covariance of the gradient vectors \mathbf{g}_{ij} , defined in (32). By the assumptions in Sec. III, these gradients have zero mean. In the sequel, we show that

$$E\{\mathbf{g}_{ij}\mathbf{g}_{mn}^\dagger\} = \begin{cases} \frac{1}{T}(\mathbf{S}_{jj} \odot \mathbf{S}_{ii}^{-1}) & (m, n) = (i, j) \\ \frac{1}{T}(\mathbf{I}_K \otimes \mathcal{T}_{m_j, m_i}) & (m, n) = (j, i) \\ \mathbf{0} & \text{o.w.} \end{cases} \quad (64)$$

The building blocks of (64) are terms of the type

$$[E\{\mathbf{g}_{ij}\mathbf{g}_{mn}^\dagger\}]_{kl} = \sum_{\alpha=1}^K \sum_{\beta=1}^K (\mathbf{I} \otimes [\mathbf{S}_{ii}^{-1}]_{k\alpha}) E\{\text{vec}\{[\bar{\mathbf{S}}_{ij}]_{\alpha k}\} \text{vec}^\dagger\{[\bar{\mathbf{S}}_{mn}]_{\beta l}\}\} (\mathbf{I} \otimes [\mathbf{S}_{mm}^{-1}]_{l\beta})^\dagger, \quad (65)$$

that relate the covariance of the gradients to the sample covariance, for any $i, j, m, n \in \{1, \dots, N\}$ and $k, l \in \{1, \dots, K\}$, as we now explain. In (65) we reformulated the k th $m_i m_j \times 1$ term of \mathbf{g}_{ij} as

$$\text{vec}\{[\mathbf{S}_{ii}^{-1}\bar{\mathbf{S}}_{ij}]_{kk}\} = \sum_{\alpha=1}^K (\mathbf{I} \otimes [\mathbf{S}_{ii}^{-1}]_{k\alpha}) \text{vec}\{[\bar{\mathbf{S}}_{ij}]_{\alpha k}\}, \quad (66)$$

which follows from applying $\text{vec}\{\cdot\}$ to

$$[\mathbf{S}_{ii}^{-1}\bar{\mathbf{S}}_{ij}]_{kk} = \sum_{\alpha=1}^K [\mathbf{S}_{ii}^{-1}]_{k\alpha} [\bar{\mathbf{S}}_{ij}]_{\alpha k} \quad (67)$$

and then (46c) in order to separate the stochastic and the deterministic terms. The equality in (67) follows from $\mathbf{S}_{ii} \in \mathcal{B}_{1_K \otimes m_i}$. Using the explicit form

$$[\bar{\mathbf{S}}_{ij}]_{\alpha k} = \frac{1}{T} \sum_{t=1}^T \mathbf{s}_i^{[\alpha]}(t) \mathbf{s}_j^{[k]\dagger}(t) \quad (68)$$

and following steps similar to those in [35, Appendix D], one obtains

$$[E\{\mathbf{g}_{ij}\mathbf{g}_{mn}^\dagger\}]_{kl} = \begin{cases} \frac{1}{T}(\mathbf{S}_{jj}^{[k,l]} \otimes [\mathbf{S}_{ii}^{-1}]_{kl}) & (m, n) = (i, j) \\ \frac{1}{T}\mathcal{T}_{m_j, m_i} \delta_{kl} & (m, n) = (j, i) \\ \mathbf{0} & \text{o.w.} \end{cases}$$

which is the block-wise form of (64). As in [35, Appendix D], the derivation is based on $E\{\mathbf{s}_i(t)\mathbf{s}_{j \neq i}^\dagger(r)\} = \mathbf{0} \forall r, t$, $E\{\mathbf{s}_i(t)\} = \mathbf{0} \forall i, t$ and sample decorrelation (in accordance with Sec. III), without resorting to any further assumptions on the distributions of the samples.

REFERENCES

- [1] 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.
- [2] J.-F. Cardoso, "Multidimensional independent component analysis," in *Proc. ICASSP*, vol. 4, Seattle, WA, May 1998, pp. 1941–1944.
- [3] A. Hyvärinen and P. O. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces," *Neural Comput.*, vol. 12, no. 7, pp. 1705–1720, Jul. 2000.
- [4] 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, Heidelberg: Springer, 2006, pp. 165–172.
- [5] Y.-O. Li, T. Adali, 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.
- [6] D. Lahat and C. Jutten, "Joint blind source separation of multidimensional components: Model and algorithm," in *Proc. EUSIPCO*, Lisbon, Portugal, Sep. 2014, pp. 1417–1421.
- [7] F. J. Theis, "Towards a general independent subspace analysis," in *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. Platt, and T. Hoffman, Eds. Cambridge, MA: MIT Press, 2007, pp. 1361–1368.
- [8] J. A. Palmer and S. Makeig, "Contrast functions for independent subspace analysis," in *Latent Variable Analysis and Signal Separation*, ser. LNCS, F. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, Eds., vol. 7191. Springer, 2012, pp. 115–122.
- [9] Z. Szabó, B. Póczos, and A. Lőrincz, "Separation theorem for independent subspace analysis and its consequences," *Pattern Recognition*, vol. 45, no. 4, pp. 1782–1791, Apr. 2012.
- [10] P. Tichavský, A. Yeredor, and Z. Koldovský, "On computation of approximate joint block-diagonalization using ordinary AJD," in *Latent Variable Analysis and Signal Separation*, ser. LNCS, F. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, Eds., vol. 7191. Springer, 2012, pp. 163–171.
- [11] L. De Lathauwer, C. Févotte, B. De Moor, and J. Vandewalle, "Jacobi algorithm for joint block diagonalization in blind identification," in *Proc. 23rd Symp. Inf. Theory in the Benelux*, Louvain-la-Neuve, Belgium, May 29–31 2002, pp. 155–162.
- [12] A. Hyvärinen and U. Köster, "FastISA: A fast fixed-point algorithm for independent subspace analysis," in *Proc. ESANN*, Bruges (Belgium), Apr. 2006, pp. 371–376.
- [13] L. De Lathauwer and D. Nion, "Decompositions of a higher-order tensor in block terms. Part III: Alternating least squares algorithms," *SIAM J. Matrix Anal. Appl.*, vol. 30, no. 3, pp. 1067–1083, 2008.
- [14] P. Tichavský and Z. Koldovský, "Algorithms for nonorthogonal approximate joint block-diagonalization," in *Proc. EUSIPCO*, Bucharest, Romania, Aug. 2012, pp. 2094–2098.
- [15] D. Lahat, J.-F. Cardoso, and H. MESSER, "Joint block diagonalization algorithms for optimal separation of multidimensional components," in *Latent Variable Analysis and Signal Separation*, ser. LNCS, F. Theis, A. Cichocki, A. Yeredor, and M. Zibulevsky, Eds., vol. 7191. Heidelberg: Springer, 2012, pp. 155–162.
- [16] M. Sørensen, I. Domanov, D. Nion, and L. De Lathauwer, "Coupled canonical polyadic decompositions and (coupled) decompositions in multilinear rank- $(L_{r,n}, L_{r,n}, 1)$ terms—part II: Algorithms," ESAT-STADIUS, KU Leuven (Leuven, Belgium), Tech. Rep. 13-144, 2014.
- [17] 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.
- [18] 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.
- [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] H. W. Gutch and F. J. Theis, "Uniqueness of linear factorizations into independent subspaces," *JMVA*, vol. 112, pp. 48–62, Nov. 2012.
- [21] 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," ESAT-STADIUS, KU Leuven (Leuven, Belgium), Tech. Rep. 13-143, 2014.
- [22] S. Ma, N. M. Correa, X.-L. Li, T. Eichele, V. D. Calhoun, and T. Adali, "Automatic identification of functional clusters in fMRI data using spatial dependence," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 12, pp. 3406–3417, Dec. 2011.

- [23] 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.
- [24] 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.
- [25] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, Dec. 1936.
- [26] T. Kim, I. Lee, and T.-W. Lee, "Independent vector analysis: Definition and algorithms," in *Proc. ACSSC*, Pacific Grove, CA, Nov. 2006, pp. 1393–1396.
- [27] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.
- [28] X.-L. Li, T. Adali, 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.
- [29] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1672–1683, Apr. 2012.
- [30] T. Adali, 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.
- [31] T. W. Anderson, *An introduction to multivariate statistical analysis*. John Wiley & Sons, 1958.
- [32] J. Chatel-Goldman, M. Congedo, and R. Phlypo, "Joint BSS as a natural analysis framework for EEG-hyperscanning," in *Proc. ICASSP*, Vancouver, Canada, May 2013, pp. 1212–1216.
- [33] S. Ma, V. D. Calhoun, R. Phlypo, and T. Adali, "Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis," *NeuroImage*, vol. 90, pp. 196–206, Apr. 2014.
- [34] R. F. Silva, S. Plis, T. Adali, and V. D. Calhoun, "Multidataset independent subspace analysis extends independent vector analysis," in *Proc. ICIP*, Paris, France, Oct. 2014, pp. 2864–2868.
- [35] 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.
- [36] 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.
- [37] S. Liu, "Matrix results on the Khatri-Rao and Tracy-Singh products," *Linear Algebra and its Applications*, vol. 289, no. 1–3, pp. 267–277, Sep. 1999.
- [38] 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.
- [39] Z. Yang, H. Lü, X. Feng, and M. Chen, "Eigenvalues estimation of block Kronecker product of positive semidefinite Hermitian matrices," in *Recent Advances in Computer Science and Information Engineering*, ser. Lecture Notes in Electrical Engineering, Z. Qian, L. Cao, W. Su, T. Wang, and H. Yang, Eds. Springer Berlin Heidelberg, 2012, vol. 126, pp. 733–739.
- [40] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, Mar. 1951.
- [41] J. Vía, M. Anderson, X.-L. Li, and T. Adali, "A maximum likelihood approach for independent vector analysis of Gaussian data sets," in *Proc. MLSP*, Beijing, China, Sep. 2011.
- [42] P. Comon, "Independent component analysis," in *Proc. Int. Signal Process. Workshop on HOS*, Chamrousse, France, Jul. 1991, pp. 111–120, keynote address. Republished in *HOS*, J.-L. Lacomme ed., Elsevier, 1992, pp. 29–38.
- [43] —, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
- [44] D.-T. Pham and P. Garat, "Blind separation of mixtures of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, Jul. 1997.
- [45] D.-T. Pham, "Joint approximate diagonalization of positive definite hermitian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136–1152, 2001.
- [46] J. R. Magnus and H. Neudecker, "The commutation matrix: Some properties and applications," *Ann. Statist.*, vol. 7, no. 2, pp. 381–394, Mar. 1979.
- [47] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adali, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4399–4410, Sep. 2014.
- [48] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, ser. Mathematics and its Applications. Chichester, West Sussex, England: Ellis Horwood Limited, 1981.
- [49] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," Oct. 2008, version 20081110. [Online]. Available: <http://www2.imm.dtu.dk/pubdb/p.php?3274>