JOINT BLIND SOURCE SEPARATION FROM SECOND-ORDER STATISTICS: NECESSARY AND SUFFICIENT IDENTIFIABILITY CONDITIONS

Javier Vía[⋆] *Matthew Anderson*[†] *Xi-Lin Li*[†] *Tülay Adalı*[†]

[⋆]Dept. of Communications Engineering. University of Cantabria. Spain
[†]Dept. of Computer Science and Electrical Engineering. University of Maryland, Baltimore County
e-mail: jvia@gtas.dicom.unican.es, matt.anderson@umbc.edu, lixilin@umbc.edu, adali@umbc.edu

ABSTRACT

This paper considers the problem of joint blind source separation (J-BSS), which appears in many practical problems such as blind deconvolution or functional magnetic resonance imaging (fMRI). In particular, we establish the necessary and sufficient conditions for the solution of the J-BSS problem by exclusively exploiting the second-order statistics (SOS) of the observations. The identifiability analysis is based on the idea of equivalently distributed sets of latent variables, that is, latent variables with covariance matrices related by means of a diagonal matrix. Interestingly, the identifiability analysis also allows us to introduce a measure of the *identifiability degree* based on Kullback-Leibler projections. This measure is clearly correlated with the performance of practical SOS-based J-BSS algorithms, which is illustrated by means of numerical examples.

Index Terms— Joint blind source separation (J-BSS), independent vector analysis (IVA), second-order statistics, identifiability.

1. INTRODUCTION

The problem of joint blind source separation (J-BSS) arises in many practical applications such as functional magnetic resonance imaging (fMRI) [1], or analysis of transformed data in multiple frequency bins. A fundamental J-BSS criterion consists in independent vector analysis (IVA) [1–3], which represents each data set as a linear combination of independent latent variables. Thus, although the data model for a single data set is identical to the ICA model [4], the existence of dependencies among the latent variables of different data sets establishes the key difference between IVA and ICA.

Interestingly, in [1] the authors have shown that the J-BSS problem can be solved by means of a practical algorithm based on the multiset extension of canonical correlation analysis (CCA) [5–8]. That is, the J-BSS problem can be solved by exclusively exploiting the second-order statistics (SOS) of the observations. However, the rigorous identifiability analysis of the J-BSS problem is still missing, and it constitutes the main goal of this work.

The main result of this paper consists in the derivation of the necessary and sufficient conditions for the solution of the J-BSS problem from the SOS of the observations. The result is based on the concept of equivalently distributed sets of latent variables, i.e., sets of latent variables with SOS (covariance matrices) related by means of a diagonal matrix (individual complex scale factors). Interestingly, the identifiability result also allows us to introduce a measure of the *identifiability degree* of a set of latent variables. This measure is based on the Kullback-Leibler divergence between zero-mean multivariate Gaussian distributions, and we show, by means of simulations, that the proposed measure is highly correlated with the practical performance of SOS-based J-BSS algorithms.

This work was supported by the Spanish Government, Ministerio de Ciencia e Innovación (MICINN), under projects MultiMIMO (TEC2007-68020-C04-02), COSIMA (TEC2010-19545-C04-03) and COMONSENS (CSD2008-00010, CONSOLIDER-INGENIO 2010 Program).

2. NOTATION AND J-BSS DATA MODEL

In this paper we use bold-faced upper case letters to denote matrices, bold-faced lower case letters for column vectors, and light-faced lower case letters for scalar quantities. Superscripts $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ denote complex conjugate, transpose and Hermitian, respectively. The notation $\mathbf{A} \in \mathbb{C}^{n\times m}$ denotes that $\mathbf{A}$ is a $n \times m$ matrix with complex entries. The absolute value and argument of the complex number a are denoted as $|a|$ and $\arg(a)$, respectively. The trace and determinant of a matrix $\mathbf{A}$ are denoted as $\mathrm{Tr}(\mathbf{A})$ and $\det(\mathbf{A})$. The diagonal matrix with vector $\mathbf{a}$ along its diagonal is denoted by $\mathrm{diag}(\mathbf{a})$, $\mathbf{I}_n$ is the identity matrix of dimension n, and $\mathbf{0}_{n\times m}$ is the $n \times m$ zero matrix. Finally, the Kronecker product is denoted by $\otimes$, E is the expectation operator, and in general $\mathbf{R}_{\mathbf{a},\mathbf{b}}$ is the cross-correlation matrix for vectors $\mathbf{a}$ and $\mathbf{b}$, i.e., $\mathbf{R}_{\mathbf{a},\mathbf{b}} = E\mathbf{a}\mathbf{b}^H$.

2.1. J-BSS Data Model

Consider M data sets (vector observations) $\mathbf{x}_m \in \mathbb{C}^{K\times 1}$ given by

$$\mathbf{x}_m = \mathbf{A}_m\mathbf{s}_m, \qquad m = 1,\ldots,M,$$

where $\mathbf{A}_m \in \mathbb{C}^{K\times K}$ are unknown non-singular mixing matrices,[1] and $\mathbf{s}_m \in \mathbb{C}^{K\times 1}$ are the unknown source vectors (or latent variables), which are modeled as zero-mean proper complex random vectors.

The goal of J-BSS consists in recovering the sources $\mathbf{s}_m$ and mixing matrices $\mathbf{A}_m$ from the observations $\mathbf{x}_m$ ($m = 1,\ldots,M$). However, unlike traditional BSS approaches, J-BSS does not only exploit the intra-set statistical relationships, but it considers the M data sets as a whole, exploiting inter-set dependencies. Specifically, the J-BSS model assumes that the entries in $\mathbf{s}_m$ are independent, but correlated with the entries in other source vectors [1]. Thus, the SOS of the sources are

$$\mathbf{R}_{\mathbf{s},\mathbf{s}} = E\mathbf{s}\mathbf{s}^H = \begin{bmatrix} \mathbf{\Lambda}_{1,1} & \mathbf{\Lambda}_{1,2} & \ldots & \mathbf{\Lambda}_{1,M} \\ \mathbf{\Lambda}_{2,1} & \mathbf{\Lambda}_{2,2} & \ldots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{\Lambda}_{M,1} & \ldots & \ldots & \mathbf{\Lambda}_{M,M} \end{bmatrix},$$

[1]We assume square mixing matrices for notational simplicity. The results in this paper can be easily extended to the over-determined case with full column-rank mixing matrices.

where $\mathbf{\Lambda}_{m,n} = E\mathbf{s}_m\mathbf{s}_n^H \in \mathbb{C}^{K\times K}$ are diagonal matrices, and $\mathbf{s} = \left[\mathbf{s}_1^T \dots \mathbf{s}_M^T\right]^T$ is defined as the concatenated source vector.

Analogously, the SOS of the observations are given by

$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = E\mathbf{x}\mathbf{x}^H = \mathbf{A}\mathbf{R}_{\mathbf{s},\mathbf{s}}\mathbf{A}^H,$$

where $\mathbf{x} = \left[\mathbf{x}_1^T \dots \mathbf{x}_M^T\right]^T$ is the concatenated observation vector, and $\mathbf{A} \in \mathbb{C}^{KM\times KM}$ is the block diagonal mixing matrix

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{0}_{K\times K} & \dots & \mathbf{0}_{K\times K} \\ \mathbf{0}_{K\times K} & \mathbf{A}_2 & \dots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{K\times K} & \dots & \dots & \mathbf{A}_M \end{bmatrix}.$$

Finally, we must note that our J-BSS data model does not assume any particular ordering of the correlation values. That is, the absolute values of the elements in the diagonal matrices $\mathbf{\Lambda}_{m,n}$ are not necessarily in decreasing order. Therefore, the data model considered in this paper is more general than that in [1].

2.2. Reordered Model

Let us introduce a permutation in the concatenated source and observation vectors $\mathbf{s}$ and $\mathbf{x}$ as follows

$$\tilde{\mathbf{x}} = \mathbf{P}\mathbf{x} = \left[\tilde{\mathbf{x}}_1^T, \dots, \tilde{\mathbf{x}}_K^T\right]^T, \quad \tilde{\mathbf{s}} = \mathbf{P}\mathbf{s} = \left[\tilde{\mathbf{s}}_1^T, \dots, \tilde{\mathbf{s}}_K^T\right]^T,$$
$$\tilde{\mathbf{x}}_k = [x_1(k), \dots, x_M(k)]^T, \quad \tilde{\mathbf{s}}_k = [s_1(k), \dots, s_M(k)]^T,$$

where $\mathbf{P} \in \mathbb{R}^{KM\times KM}$ is a permutation matrix, $x_m(k)$ denotes the k-th element in $\mathbf{x}_m$, and $s_m(k)$ is defined in the same way. With these definitions, the data model can be rewritten as

$$\mathbf{R}_{\tilde{\mathbf{x}},\tilde{\mathbf{x}}} = \tilde{\mathbf{A}}\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\tilde{\mathbf{A}}^H,$$

where

$$\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}} = \mathbf{P}\mathbf{R}_{\mathbf{s},\mathbf{s}}\mathbf{P}^T = \begin{bmatrix} \mathbf{R}_{\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_1} & \mathbf{0}_{M\times M} & \dots & \mathbf{0}_{M\times M} \\ \mathbf{0}_{M\times M} & \mathbf{R}_{\tilde{\mathbf{s}}_2,\tilde{\mathbf{s}}_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{M\times M} & \dots & \dots & \mathbf{R}_{\tilde{\mathbf{s}}_K,\tilde{\mathbf{s}}_K} \end{bmatrix},$$

contains the correlation matrices associated to each set of latent variables, and

$$\tilde{\mathbf{A}} = \mathbf{P}\mathbf{A}\mathbf{P}^T = \begin{bmatrix} \tilde{\mathbf{A}}_{1,1} & \dots & \tilde{\mathbf{A}}_{1,K} \\ \vdots & \ddots & \vdots \\ \tilde{\mathbf{A}}_{K,1} & \dots & \tilde{\mathbf{A}}_{K,K} \end{bmatrix},$$

is a global mixing matrix with diagonal blocks given by

$$\tilde{\mathbf{A}}_{i,j} = \text{diag}\left([a_1(i,j), \dots, a_M(i,j)]\right),$$

where $a_m(i,j)$ denotes the entry in the i-th row and j-th column of $\mathbf{A}_m$.

3. IDENTIFIABILITY ANALYSIS

In order to recover the sources $\mathbf{s}$ and mixing matrix $\mathbf{A}$ from the SOS of the observations $\mathbf{x}$, we need to exploit the particular structure of the matrices $\mathbf{A}$ and $\mathbf{R}_{\mathbf{s},\mathbf{s}}$ (equivalently $\tilde{\mathbf{A}}$ and $\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}$). That is, we are looking for a pair of matrices $\hat{\mathbf{A}}$, $\hat{\mathbf{R}}_{\mathbf{s},\mathbf{s}}$ such that

$$\mathbf{R}_{\mathbf{x},\mathbf{x}} = \mathbf{A}\mathbf{R}_{\mathbf{s},\mathbf{s}}\mathbf{A}^H = \hat{\mathbf{A}}\hat{\mathbf{R}}_{\mathbf{s},\mathbf{s}}\hat{\mathbf{A}}^H. \tag{1}$$

3.1. Trivial Ambiguities

Before proceeding, we have to take into account two trivial ambiguities inherent to the J-BSS model:

- Individual non-null scale factors for each of the source vectors $\mathbf{s}_m$ and mixing matrices $\mathbf{A}_m$.
- A common permutation matrix for all the source vectors $\mathbf{s}_m$ and mixing matrices $\mathbf{A}_m$.

In other words, for all permutation matrices $\mathbf{\Pi} \in \mathbb{R}^{K\times K}$ and non-singular diagonal matrices $\mathbf{\Phi}_m \in \mathbb{C}^{K\times K}$, we have that the set of matrices

$$\hat{\mathbf{A}}_m = \mathbf{A}_m\mathbf{\Phi}_m\mathbf{\Pi}, \qquad \hat{\mathbf{\Lambda}}_{m,n} = \mathbf{\Pi}^T\mathbf{\Phi}_m^{-1}\mathbf{\Lambda}_{m,n}\mathbf{\Phi}_n^{-1}\mathbf{\Pi},$$

are a solution of the J-BSS model. Thus, without loss of generality, we can assume unit-variance sources ($\mathbf{\Lambda}_{m,m} = \mathbf{I}_K$), and using (1) we have

$$\hat{\mathbf{A}}_m\hat{\mathbf{A}}_m^H = \mathbf{A}_m\mathbf{A}_m^H, \qquad m = 1, \dots, M,$$

which implies

$$\hat{\mathbf{A}}_m = \mathbf{A}_m\mathbf{Q}_m, \qquad m = 1, \dots, M,$$

and

$$\mathbf{\Lambda}_{m,n} = \mathbf{Q}_m\hat{\mathbf{\Lambda}}_{m,n}\mathbf{Q}_n^H, \qquad m,n = 1, \dots, M, \tag{2}$$

where $\mathbf{Q}_m \in \mathbb{C}^{K\times K}$ are unitary matrices.

3.2. Identifiability Conditions

In this subsection, we establish the necessary and sufficient conditions for the identifiability of the J-BSS model from SOS. In order to simplify the presentation and proof of the main result, we will consider non-null correlation values,[2]

$$\lambda_{m,n}(k) \neq 0, \qquad \forall m,n,k,$$

where $\lambda_{m,n}(k)$ is the k-th element in the diagonal of $\mathbf{\Lambda}_{m,n}$.

Let us start by introducing the definition of equivalently distributed sets of latent variables.

Definition 1 *Two sets of latent variables $\tilde{\mathbf{s}}_k$, $\tilde{\mathbf{s}}_l$ are* **equivalently distributed** *if and only if their correlation matrices are related as*

$$\mathbf{R}_{\tilde{\mathbf{s}}_k,\tilde{\mathbf{s}}_k} = \mathbf{\Psi}\mathbf{R}_{\tilde{\mathbf{s}}_l,\tilde{\mathbf{s}}_l}\mathbf{\Psi}^H,$$

where $\mathbf{\Psi} \in \mathbb{C}^{M\times M}$ is a (non-singular) diagonal matrix.

The above definition is clearly related to the trivial ambiguity consisting in individual scale factors of the mixing matrices and latent variables, and it plays a key role in the identifiability analysis. In particular, the necessary and sufficient identifiability conditions are stated in the following theorem, whose proof is relegated to the next subsection.

Theorem 1 (J-BSS Identifiability Conditions) *Given the J-BSS model $\mathbf{x}_m = \mathbf{A}_m\mathbf{s}_m$ ($m = 1, \dots, M$), the sources and mixing matrices can be recovered from the SOS of the observations up to the trivial ambiguities and a non-trivial common mixture among the equivalently distributed sets of latent variables.*

[2] Similar identifiability conditions can be found in the case of vanishing correlation values. However, this comes at the expense of a more complicated notation.

Theorem 1 establishes very mild identifiability conditions which, as we will see in the simulations section, are satisfied with probability one for randomly generated correlation matrices $\mathbf{R}_{\tilde{\mathbf{s}}_k,\tilde{\mathbf{s}}_k}$. Moreover, as a direct consequence of Theorem 1, we can say that (all) the sources can be recovered using the SOS when no two sets of latent variables are equivalently distributed. Finally, we must point out that the identifiability conditions in this paper are less restrictive than those of the particular J-BSS algorithm proposed in [1], even though the model considered here is more general.

3.3. Proof of Theorem 1

Let us start by using (2) to write

$$\mathbf{\Sigma}_{m,n} = \mathbf{Q}_m \hat{\mathbf{\Sigma}}_{m,n} \mathbf{Q}_m^H, \qquad m,n = 1,\ldots,M, \tag{3}$$

where $\mathbf{\Sigma}_{m,n} = \mathbf{\Lambda}_{m,n}\mathbf{\Lambda}_{m,n}^H$ and $\hat{\mathbf{\Sigma}}_{m,n} = \hat{\mathbf{\Lambda}}_{m,n}\hat{\mathbf{\Lambda}}_{m,n}^H$. Thus, the above equation represents the eigenvalue decomposition (EVD) of the real and diagonal matrix $\mathbf{\Sigma}_{m,n}$, and the orthogonal eigenvectors in $\mathbf{Q}_m$ are unique up to complex scale factors, a permutation, and rotations involving the eigenvectors associated with equal eigenvalues. Let us now consider that there are more than one non-null entry in the i-th column of $\mathbf{Q}_m$, i.e., that there exists a pair of indices k,l such that

$$q_m(k,i) \neq 0, \qquad q_m(l,i) \neq 0, \tag{4}$$

where $q_m(k,i)$ is the element in the k-th row and i-th column of $\mathbf{Q}_m$. Then, taking (3) into account, we can see that (4) implies

$$|\lambda_{m,n}(k)| = |\lambda_{m,n}(l)| = |\hat{\lambda}_{m,n}(k)| = |\hat{\lambda}_{m,n}(l)|, \qquad \forall n,$$

where $\lambda_{m,n}(k)$ (respectively $\hat{\lambda}_{m,n}(k)$) is the k-th element in the diagonal of $\mathbf{\Lambda}_{m,n}$ (resp. $\hat{\mathbf{\Lambda}}_{m,n}$). Furthermore, taking into account $|\lambda_{m,n}(k)| = |\lambda_{m,n}(l)| \neq 0$, we can easily deduce from (2) that (4) implies

$$q_n(k,i) \neq 0, \qquad q_n(l,i) \neq 0, \qquad \forall n,$$

which directly results in

$$|\lambda_{m,n}(k)| = |\lambda_{m,n}(l)| = |\hat{\lambda}_{m,n}(k)| = |\hat{\lambda}_{m,n}(l)|, \qquad \forall m,n. \tag{5}$$

Now, using (2), we can write

$$\mathbf{\Lambda}_{m,n}\mathbf{\Lambda}_{n,p} = \mathbf{Q}_m \hat{\mathbf{\Lambda}}_{m,n}\hat{\mathbf{\Lambda}}_{n,p}\mathbf{Q}_p^H, \qquad m,n,p = 1,\ldots,M,$$

and combining this equation with $\mathbf{\Lambda}_{m,p} = \mathbf{Q}_m\hat{\mathbf{\Lambda}}_{m,p}\mathbf{Q}_p^H$ we get, for $m,n,p = 1,\ldots,M$,

$$\mathbf{\Lambda}_{m,n}\mathbf{\Lambda}_{n,p} + \mathbf{\Lambda}_{m,p} = \mathbf{Q}_m\left(\hat{\mathbf{\Lambda}}_{m,n}\hat{\mathbf{\Lambda}}_{n,p} + \hat{\mathbf{\Lambda}}_{m,p}\right)\mathbf{Q}_p^H,$$
$$\mathbf{\Lambda}_{m,n}\mathbf{\Lambda}_{n,p} + j\mathbf{\Lambda}_{m,p} = \mathbf{Q}_m\left(\hat{\mathbf{\Lambda}}_{m,n}\hat{\mathbf{\Lambda}}_{n,p} + j\hat{\mathbf{\Lambda}}_{m,p}\right)\mathbf{Q}_p^H.$$

Thus, from the above equations we can deduce that (4) implies (for all m,n,p)

$$|\lambda_{m,n}(k)\lambda_{n,p}(k) + \lambda_{m,p}(k)| = |\lambda_{m,n}(l)\lambda_{n,p}(l) + \lambda_{m,p}(l)|,$$
$$|\lambda_{m,n}(k)\lambda_{n,p}(k) + j\lambda_{m,p}(k)| = |\lambda_{m,n}(l)\lambda_{n,p}(l) + j\lambda_{m,p}(l)|,$$

which combined with (5) yields

$$\lambda_{m,n}(k)\lambda_{n,p}(k)\lambda_{m,p}^*(k) = \lambda_{m,n}(l)\lambda_{n,p}(l)\lambda_{m,p}^*(l),$$

or equivalently

$$\lambda_{n,p}(k) = e^{j\theta_{m,n}}\lambda_{n,p}(l)e^{-j\theta_{m,p}}, \qquad m,n,p = 1,\ldots,M,$$

where $\theta_{m,n} = \arg(\lambda_{m,n}(l)) - \arg(\lambda_{m,n}(k))$.

Finally, taking into account the assumption of unit variance sources, the above equation can be rewritten in a more compact way as

$$\mathbf{R}_{\tilde{\mathbf{s}}_k,\tilde{\mathbf{s}}_k} = \mathbf{\Psi}\mathbf{R}_{\tilde{\mathbf{s}}_l,\tilde{\mathbf{s}}_l}\mathbf{\Psi}^H,$$

where, for $m = 1,\ldots,M$,

$$\mathbf{\Psi} = e^{j\theta_m}\text{diag}\left(\left[e^{j\theta_{m,1}},\ldots,e^{j\theta_{m,M}}\right]\right),$$

and $\theta_m \in \mathbb{R}$ represents an arbitrary change of phase.

Summarizing, we have proven that the existence of non-trivial ambiguities relating $\tilde{\mathbf{s}}_k$ and $\tilde{\mathbf{s}}_l$ implies that these sets of latent variables are equivalently distributed. The proof of the converse is straightforward. Consider P sets of equivalently distributed latent variables. Then, the associated matrix $\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}$ can be written as

$$\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}} = \tilde{\mathbf{\Psi}}\left(\mathbf{I}_P \otimes \mathbf{R}_{\tilde{\mathbf{s}}_1,\tilde{\mathbf{s}}_1}\right)\tilde{\mathbf{\Psi}}^H,$$

with

$$\tilde{\mathbf{\Psi}} = \begin{bmatrix} \mathbf{I}_M & \mathbf{0}_{M\times M} & \ldots & \mathbf{0}_{M\times M} \\ \mathbf{0}_{M\times M} & \mathbf{\Psi}_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0}_{M\times M} & \ldots & \ldots & \mathbf{\Psi}_P \end{bmatrix},$$

and $\mathbf{\Psi}_2,\ldots,\mathbf{\Psi}_P$ diagonal complex matrices. Therefore, introducing a mixing matrix

$$\tilde{\mathbf{A}} = (\mathbf{B}\otimes\mathbf{I}_M)\tilde{\mathbf{\Psi}}^{-1},$$

with $\mathbf{B} \in \mathbb{C}^{P\times P}$ a non-singular matrix, it is easy to see that the mixture $\tilde{\mathbf{A}}\mathbf{R}_{\tilde{\mathbf{s}},\tilde{\mathbf{s}}}\tilde{\mathbf{A}}^H$ results in a block-diagonal matrix satisfying the J-BSS model. This means that we can not recover the sets of equivalently distributed latent variables without exploiting some additional properties of the sources or the mixing matrix.

4. SIMULATION EXAMPLE

The main result of the paper is illustrated in this section by means of some numerical examples. In particular, we have considered circular complex Gaussian sources, which have been generated as $\tilde{\mathbf{s}}_k = \mathbf{B}_k\mathbf{u}_k$ ($k = 1,\ldots,K$), where the entries in the vectors $\mathbf{u}_k$ and matrices $\mathbf{B}_k$ are i.i.d. circular complex Gaussian with zero mean and unit variance. That is, the correlation matrices $\mathbf{R}_{\tilde{\mathbf{s}}_k,\tilde{\mathbf{s}}_k} \in \mathbb{C}^{M\times M}$ for the sets of latent variables are complex Wishart. Analogously, the entries in the mixing matrices $\mathbf{A}_m \in \mathbb{C}^{K\times K}$ ($m = 1,\ldots,M$) are i.i.d. circular complex Gaussian with zero mean and unit variance.

In all the experiments we have used four-dimensional vectors $K = 4$, and we have considered two different numbers of data sets ($M = 3$ and $M = 10$). The J-BSS is performed by means of a new algorithm based on the maximum likelihood estimation of the mixing matrices $\mathbf{A}_m$ in the case of Gaussian data.[3] The J-BSS performance measure is based on the residual mixture matrices

$$\mathbf{E}_m = \hat{\mathbf{A}}_m^{-1}\mathbf{A}_m, \qquad m = 1,\ldots,M,$$

where $\hat{\mathbf{A}}_m$ and $\mathbf{A}_m$ denote the estimated and actual mixing matrices. In particular, after solving the permutation ambiguity, the residual mixture measure between the k-th and l-th sets of latent variables is defined as

$$J_{k,l} = \frac{1}{M}\sum_{m=1}^{M}\frac{|\mathbf{E}_m(k,l)|^2}{|\mathbf{E}_m(k,k)|^2},$$

[3] The details of the algorithm are not included here due to the lack of space, but they will be reported in a future paper.

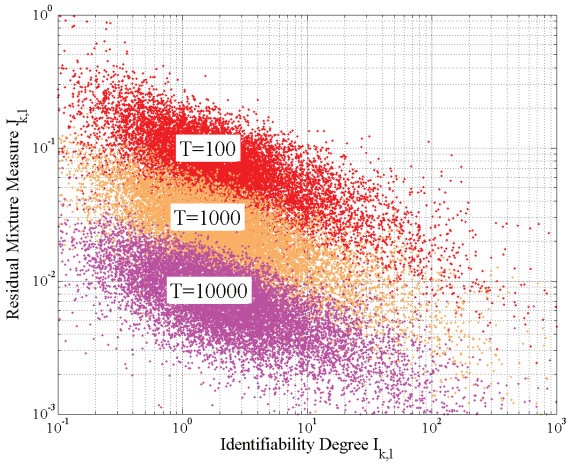

Fig. 1. J-BSS example. $M = 3$ data sets with dimension $K = 4$.

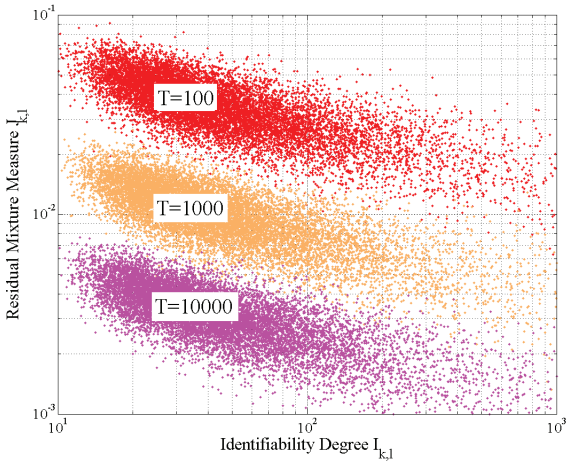

Fig. 2. J-BSS example. $M = 10$ data sets with dimension $K = 4$.

where $\mathbf{E}_m(k,l)$ is the entry in the k-th row and l-th column of $\mathbf{E}_m$. That is, $J_{k,l}$ represents the *interference* of $\tilde{\mathbf{s}}_l$ in the estimated version of $\tilde{\mathbf{s}}_k$.

On the other hand, the *identifiability degree* is measured by means of the Kullback-Leibler (KL) divergence [9] between pairs of sets of latent variables. In particular, we define

$$I_{k,l} = \underset{\mathbf{\Psi}=\text{diag}([\psi_1,\ldots,\psi_M])}{\text{minimum}} D_{\text{KL}}(\mathbf{R}_{\tilde{\mathbf{s}}_l,\tilde{\mathbf{s}}_l} | \mathbf{\Psi}\mathbf{R}_{\tilde{\mathbf{s}}_k,\tilde{\mathbf{s}}_k}\mathbf{\Psi}),$$

where

$$D_{\text{KL}}(\mathbf{R}_0|\mathbf{R}_1) = \log\frac{\det(\mathbf{R}_1)}{\det(\mathbf{R}_0)} + \text{Tr}\left(\mathbf{R}_1^{-1}\mathbf{R}_0\right) - M,$$

is the KL divergence between zero-mean circular complex Gaussian distributions with covariance matrices $\mathbf{R}_0 \in \mathbb{C}^{M\times M}$ and $\mathbf{R}_1 \in \mathbb{C}^{M\times M}$. This measure can be seen as the *distance* between $\mathbf{R}_{\tilde{\mathbf{s}}_l,\tilde{\mathbf{s}}_l}$ and the closest (in the KL sense) $\mathbf{\Psi}\mathbf{R}_{\tilde{\mathbf{s}}_k,\tilde{\mathbf{s}}_k}\mathbf{\Psi}$. Obviously, $I_{k,l}$ is always nonnegative, and it is zero if and only if the sets of latent variables $\tilde{\mathbf{s}}_k$, $\tilde{\mathbf{s}}_l$ are equivalently distributed.

Figures 1 and 2 show the simulation results for $M = 3$ and $M = 10$, different numbers (T) of available observation vectors, and 1000 independent experiments. As we can see, there exists a clear correlation between the J-BSS performance $J_{k,l}$ and the *identifiability degree* measured by $I_{k,l}$. That is, the identifiability result in this paper also provides a measure of the expected J-BSS accuracy. Furthermore, as we could expect, the performance of the J-BSS algorithm is improved by increasing the number of observations T and/or the number of data sets M. Obviously, increasing T results in more accurate estimates of the SOS, whereas increasing M makes the set of latent variables *less* equivalently distributed, which is corroborated by the different ranges of $I_{k,l}$ values in the figures.

5. CONCLUSIONS

In this paper, we have derived the necessary and sufficient conditions for joint blind source separation (J-BSS) from the second-order statistics (SOS) of the observations. The main result in the paper is based on the definition of equivalently distributed sets of latent variables, which play a key role in the J-BSS identifiability analysis. Moreover, the identifiability results also allow us to propose a measure of the expected performance of practical J-BSS algorithms, which has been illustrated by means of numerical examples.

6. REFERENCES

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

[2] 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*, Lecture Notes in Computer Science, pp. 165–172. Springer Berlin / Heidelberg, 2006.

[3] X.-L. Li, T. Adalı, and M. Anderson, "Joint blind source separation by generalizad joint diagonalization of cumulant matrices," *Signal Processing* (submitted for publication).

[4] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Wiley Interscience, 2001.

[5] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, pp. 321–377, 1936.

[6] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[7] J. Vía, I. Santamaría, and J. Pérez, "A learning algorithm for adaptive canonical correlation analysis of several data sets," *Neural Networks*, vol. 20, no. 1, pp. 139–152, Jan. 2007.

[8] N. Correa, T. Adalı, Y.-O. Li, and V. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, July 2010.

[9] T.M. Cover and J.A. Thomas, *Elements of information theory*, Wiley-Interscience, New York, NY, USA, 1991.