

The Extended “Sequentially Drilled” Joint Congruence Transformation and its Application in Gaussian Independent Vector Analysis

Amir Weiss¹, Arie Yeredor¹, Sher Ali Cheema², and Martin Haardt²

Abstract—Independent Vector Analysis (IVA) has emerged in recent years as an extension of Independent Component Analysis (ICA) into multiple sets of mixtures, where the source signals in each set are independent, but may depend on source signals in the other sets. In a semi-blind IVA (or ICA) framework, information regarding the probability distributions of the sources may be available, giving rise to Maximum Likelihood (ML) separation. In recent work we have shown that under the multivariate Gaussian model, with arbitrary temporal covariance matrices (stationary or non-stationary) of the source signals, ML separation requires the solution of a “Sequentially Drilled” Joint Congruence (SeD-JoCo) transformation of a set of matrices, which is reminiscent of (but different from) classical joint diagonalization. In this paper we extend our results to the IVA problem, showing how the ML solution for the Gaussian model (with arbitrary covariance and cross-covariance matrices) takes the form of an extended SeDJoCo problem. We formulate the extended problem, derive a condition for the existence of a solution, and propose two iterative solution algorithms. In addition, we derive the induced Cramér-Rao Lower Bound (iCRLB) on the resulting Interference-to-Source Ratios (ISR) matrices, and demonstrate by simulation how the ML separation obtained by solving the extended SeD-JoCo problem indeed attains the iCRLB (asymptotically), as opposed to other separation approaches, which cannot exploit prior knowledge regarding the sources distributions.

Index Terms—Joint blind source separation, independent vector analysis, maximum likelihood, SeDJoCo.

I. INTRODUCTION

Joint matrix transformations and tensor decompositions are important fundamental algebraic tools in a broad range of signal processing fields, such as Blind Source Separation (BSS, [1]–[4]), Independent Vector Analysis (IVA, [5]–[7]), data mining ([8]–[11]) and Multi-User Multiple-Input Multiple-Output (MU-MIMO) systems in wireless communications ([12]–[15]). In many data analysis problems, special internal structures can be revealed by applying some transformations or decompositions to sets of matrices (or to tensors) constructed from the available data.

One common example is the use of Approximate Joint Diagonalization (AJD), which is instrumental in the context of BSS and independent component analysis (ICA). In many separation algorithms (e.g., JADE [16], SOBI [17], ACMA

[18]), the demixing-matrix is estimated as the matrix which best “jointly diagonalizes”, by some chosen criterion, a set of N “target-matrices” $\mathbf{Q}_1, \dots, \mathbf{Q}_N$, constructed (in some way) from the observed mixtures (e.g., a set of sample correlation matrices at different lags). The procedure of AJD of such a set of N matrices, each of dimensions $K \times K$, can be viewed as a symmetric canonical decomposition (e.g., [19]), representing (or approximating) the respective three-way $K \times K \times N$ tensor as the sum of K rank-1 tensors.

A. From SeDJoCo to Extended SeDJoCo

A particular case of this paradigm is the “Sequentially Drilled” Joint Congruence (SeDJoCo) transformation, also termed “Hybrid Exact-Approximate joint Diagonalization (HEAD)” in ([20], [21]). SeDJoCo essentially prescribes the likelihood equations for Maximum Likelihood (ML) estimation in the semi-blind separation scenario under a multivariate Gaussian model. More specifically, consider the classic linear, static, memoryless mixture model

$$\mathbf{X} = \mathbf{A}\mathbf{S}, \quad (1)$$

where $\mathbf{A} \in \mathbb{R}^{K \times K}$ is the unknown, deterministic (invertible) mixing-matrix, $\mathbf{S} = [\mathbf{s}_1 \dots \mathbf{s}_K]^T \in \mathbb{R}^{K \times T}$ is the sources matrix of K statistically independent source signals ($\mathbf{s}_1, \dots, \mathbf{s}_K \in \mathbb{R}^T$) each of length T (where $(\cdot)^T$ denotes the transpose), and $\mathbf{X} \in \mathbb{R}^{K \times T}$ is the observation matrix which contains the observed mixture signals. It has been shown (e.g., [21], [22] and [23] (chapter 7)), that when the source signals are zero-mean Gaussian, each with a known temporal covariance matrix $\mathbf{C}_k \triangleq E[\mathbf{s}_k \mathbf{s}_k^T] \in \mathbb{R}^{T \times T}$ (all distinct from each other), the ML estimate $\hat{\mathbf{B}}$ of $\mathbf{B} \triangleq \mathbf{A}^{-1}$ can be obtained (up to an inevitable sign ambiguity) as follows:

- I. Construct $N = K^1$ symmetric target-matrices as

$$\mathbf{Q}_k \triangleq \frac{1}{T} \mathbf{X} \mathbf{C}_k^{-1} \mathbf{X}^T \in \mathbb{R}^{K \times K}, \quad \forall k \in \{1, \dots, K\}. \quad (2)$$

- II. Find a matrix $\hat{\mathbf{B}} \in \mathbb{R}^{K \times K}$ that satisfies the following set of K^2 equations (K vector equations, K elements each)

$$\hat{\mathbf{B}} \mathbf{Q}_k \hat{\mathbf{B}}^T \mathbf{e}_k = \mathbf{e}_k, \quad \forall k \in \{1, \dots, K\}, \quad (3)$$

where the pinning vector \mathbf{e}_k denotes the k -th column of the $K \times K$ identity matrix.

¹Note that unlike the general (and heuristic) AJD approach, in SeDJoCo the number of target-matrices equals the number of sources

¹ Authors with Tel-Aviv University, School of Electrical Engineering, P. O. Box 39040, Tel-Aviv 69978, Israel, e-mail: amirwei2@mail.tau.ac.il, arie@eng.tau.ac.il

² Authors with Ilmenau University of Technology, Communications Research Laboratory, P. O. Box 10 05 65, D-98684 Ilmenau, Germany, e-mail: {sher-ali.cheema, martin.haardt}@tu-ilmenau.de, phone: +49 (3677) 69-2613, fax: +49 (3677) 69-1195, WWW: http://www.tu-ilmenau.de/crl.

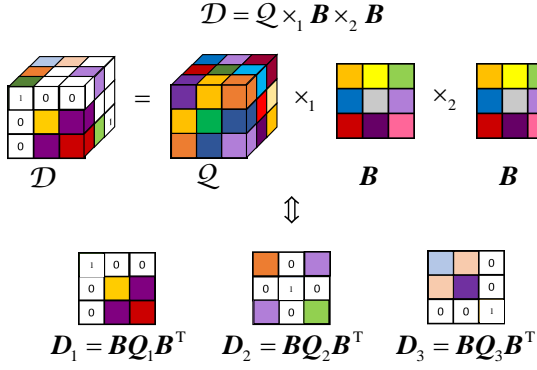


Fig. 1: Illustration of SeDJoCo as a tensor decomposition or, equivalently, as a joint matrix transformation. Here, \times_n denotes the n -mode product of a tensor with a matrix.

Condition (3) is equivalent to requiring that the matrices $D_k \triangleq \widehat{B} Q_k \widehat{B}^T$ each satisfy $D_k e_k = e_k$ (for $k = 1, \dots, K$), meaning that the vector e_k is an eigenvector of the matrix D_k with eigenvalue 1. We term this structure a “drilled” structure, hence the name of this transformation. Notice that $e_k^T D_k = e_k^T$ also holds, by the symmetry of all D_k . This special “drilled” structure of SeDJoCo, and its interpretation as a tensor decomposition, are illustrated in Fig. 1 for the case of $K = 3$. Here D_1 , D_2 , and D_3 are treated as the first, second, and third frontal slices of the tensor \mathcal{D} , respectively. The target matrices Q_1 , Q_2 , and Q_3 are associated with \mathcal{Q} in the same manner.

Another context in which the SeDJoCo solution is useful is the MU-MIMO broadcast channel Coordinated Beamforming (CBF) [13], [24], when a transmitter with K antennas transmits data to $N \leq K$ users, each with K antennas. In order to find the beamformer which perfectly eliminates all inter-users interference, a very similar (complex-valued) formulation of the SeDJoCo equations is obtained. In this case the target-matrices are defined as

$$Q_k = H_k^H H_k \in \mathbb{C}^{K \times K}, \quad \forall k \in \{1, \dots, K\}, \quad (4)$$

where $(\cdot)^H$ denotes the conjugate transpose, the matrices $H_k \in \mathbb{C}^{K \times K}$ denote the (flat fading) channel coefficients from each of the K transmit antennas to each of the K receive antennas of the k -th user ($k \in \{1, \dots, K\}$). The solution matrix \widehat{B} in this context contains the desired transmission beamforming coefficients, such that its k -th row contains the coefficients for transmission to the k -th user (see [24] for a detailed derivation in this context).

In recent years there has been a growing interest in the generalization of a single dataset BSS problem to multiple datasets BSS problem, termed Joint Blind Source Separation (JBSS, [25], [26] and references therein). Examples of applications that motivate the interest in JBSS are the analysis of multi-subject datasets of electroencephalography (EEG) data ([5], [27]) and functional magnetic resonance imaging (fMRI, [28]). A particular case of JBSS is the extension of the ICA model, termed IVA, which is formulated as follows. Consider M datasets of mixtures

$$X^{(m)} = A^{(m)} S^{(m)}, \quad \forall m \in \{1, \dots, M\}, \quad (5)$$

where $S^{(m)} = [s_1^{(m)} \dots s_K^{(m)}]^T \in \mathbb{R}^{K \times T}$ denotes a matrix of K source signals of length T (for all $m \in \{1, \dots, M\}$), belonging to the m -th dataset out of M such sets. In each dataset the sources are mixed with an unknown (deterministic) respective mixing-matrix $A^{(m)} \in \mathbb{R}^{K \times K}$, and the observed mixture signals are given by $X^{(m)} \in \mathbb{R}^{K \times T}$. Based on the observed mixtures datasets $\{X^{(m)}\}_{m=1}^M$, it is desired to estimate all M mixing-matrices and thereby recover the source signals. In the same manner as in the standard ICA model, in IVA, too, the sources within each dataset are assumed to be mutually statistically independent. Clearly, IVA amounts to M independent standard ICA problems when no statistical dependence between source signals across different datasets exists. However, in IVA statistical dependence between respective sources from different datasets is considered, i.e., the vector $s_k^{(m_1)}$ may depend on the vector $s_k^{(m_2)}$ (for all $m_1, m_2 \in \{1, \dots, M\}$ and all $k \in \{1, \dots, K\}$), but any two vectors $s_{k_1}^{(m_1)}$ and $s_{k_2}^{(m_2)}$ are statistically independent when $k_1 \neq k_2$ for any $m_1, m_2 \in \{1, \dots, M\}$. One example which is suitable for this model is group fMRI data analysis, where coherence between estimates of the source signals across different subjects (i.e., datasets) is exploited for post-analysis of the data, e.g., for group level inference and for the study of inter-subject variability [29]. Another example is the separation of mixtures of color images [7]. Suppose that we are given a set of $K = 2$ linear mixtures of two color images, where each image consists of three color layers (Red, Green and Blue), and the respective layers from each image are mixed separately so as to form the respective layers of the mixed images. So there are $M = 3$ mixtures sets, one for each color layer. The two mixed images are independent, but the color layers in each image are usually strongly correlated, giving rise to dependence between sets while maintaining independence within sets.

As we shall show immediately, it turns out that in the zero-mean Gaussian model the resulting likelihood equations for obtaining the ML estimates of the matrices $\{B^{(m)} \triangleq A^{(m)-1}\}_{m=1}^M$ in the IVA problem require a solution of what can be regarded as an extension to the SeDJoCo problem, which we term an “extended” SeDJoCo problem.

B. Extended SeDJoCo as the Likelihood Equations

In order to simplify the exposition, we introduce an equivalent formulation for the IVA model (5). Define the block diagonal matrix $\bar{A} \triangleq \text{Bdiag}(A^{(1)}, \dots, A^{(M)}) \in \mathbb{R}^{KM \times KM}$, where the $\text{Bdiag}(\cdot)$ operator creates a block-diagonal matrix from its square matrix arguments. Additionally, define the matrices $\bar{S} \triangleq [S^{(1)T} \dots S^{(M)T}]^T \in \mathbb{R}^{KM \times T}$ and $\bar{X} \triangleq [X^{(1)T} \dots X^{(M)T}]^T \in \mathbb{R}^{KM \times T}$. Model (5) can now be more compactly expressed as

$$\bar{X} = \bar{A} \bar{S}. \quad (6)$$

Left-multiplying by $\bar{B} \triangleq \bar{A}^{-1}$ and applying the $\text{vec}(\cdot)$ operator (which concatenates the columns of an $M \times N$ matrix into an $MN \times 1$ vector) we get

$$\text{vec}(\bar{S}) = (\mathbf{I}_T \otimes \bar{B}) \text{vec}(\bar{X}) \in \mathbb{R}^{KMT \times 1}, \quad (7)$$

where \mathbf{I}_T is the identity matrix of dimension T and \otimes denotes the Kronecker product.

At this point we recall that the term “blind” usually implies that no information is available regarding the sources, except for their mutual independence within each dataset (hence, the term IVA). In a “semi-blind” scenario, more *a-priori* structural or statistical information about the sources might be available. This paper addresses the semi-blind scenario with *a-priori* knowledge regarding the joint distribution of the sources. In particular, we assume that the source signals are zero-mean Gaussian with (known) temporal covariance matrices $\mathbf{C}_k^{(m_1, m_2)} \triangleq E[\mathbf{s}_k^{(m_1)} \mathbf{s}_k^{(m_2)T}] \in \mathbb{R}^{T \times T}$, namely $\mathbf{C}_k^{(m_1, m_2)}$ is the temporal covariance matrix between the k -th source at the m_1 -th set and the k -th source at the m_2 -th set. Admittedly, such a scenario may seem too far-fetched in practice. We provide some possible justifications thereto (including a description of a practical example where such prior knowledge may be available) in the next subsection. However, to proceed with the exposition, assume for now that such prior knowledge is indeed available.

Once the distribution of the sources is known (denoted by $p_{\bar{\mathbf{S}}}(\bar{\mathbf{S}})$), the parameterized probability density of the observed mixtures can be expressed as

$$p_{\bar{\mathbf{X}}}(\bar{\mathbf{X}}; \bar{\mathbf{B}}) = |\det(\mathbf{I}_T \otimes \bar{\mathbf{B}})| p_{\bar{\mathbf{S}}}((\mathbf{I}_T \otimes \bar{\mathbf{B}}) \text{vec}(\bar{\mathbf{X}})) \\ = |\det \bar{\mathbf{B}}|^T p_{\bar{\mathbf{S}}}((\mathbf{I}_T \otimes \bar{\mathbf{B}}) \text{vec}(\bar{\mathbf{X}})). \quad (8)$$

For the explicit expression of the likelihood function of $\bar{\mathbf{B}}$ we must obtain an explicit form of $p_{\bar{\mathbf{S}}}(\bar{\mathbf{S}})$. Fortunately, things can be simplified by exploiting the statistical independence between all Source Component Vectors (SCVs), defined as

$$\bar{\mathbf{s}}_k \triangleq \text{vec}(\mathbf{S}_k^T), \quad \forall k \in \{1, \dots, K\}, \quad (9)$$

where \mathbf{S}_k is the k -th source component matrix, defined as

$$\mathbf{S}_k \triangleq [\mathbf{s}_k^{(1)} \dots \mathbf{s}_k^{(M)}]^T \in \mathbb{R}^{M \times T}. \quad (10)$$

The covariance matrix of each SCV is given by

$$\bar{\mathbf{C}}_k \triangleq E[\bar{\mathbf{s}}_k \bar{\mathbf{s}}_k^T] = \begin{bmatrix} \mathbf{C}_k^{(1,1)} & \dots & \mathbf{C}_k^{(1,M)} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_k^{(M,1)} & \dots & \mathbf{C}_k^{(M,M)} \end{bmatrix} \in \mathbb{R}^{MT \times MT}, \quad (11)$$

and we denote the respective block-partition of its inverse as

$$\bar{\mathbf{C}}_k^{-1} \triangleq \begin{bmatrix} \mathbf{P}_k^{(1,1)} & \dots & \mathbf{P}_k^{(1,M)} \\ \vdots & \ddots & \vdots \\ \mathbf{P}_k^{(M,1)} & \dots & \mathbf{P}_k^{(M,M)} \end{bmatrix} \triangleq \bar{\mathbf{P}}_k, \quad (12)$$

where $\mathbf{P}_k^{(m_1, m_2)} \in \mathbb{R}^{T \times T}$, to be used below. Using the Gaussian distribution of the sources we have

$$p_{\bar{\mathbf{S}}}(\bar{\mathbf{S}}) = \prod_{k=1}^K p_{\bar{\mathbf{s}}_k}(\bar{\mathbf{s}}_k) \\ = \prod_{k=1}^K (2\pi)^{-\frac{MT}{2}} |\det \bar{\mathbf{C}}_k|^{-\frac{1}{2}} \exp\left(-\frac{1}{2} \bar{\mathbf{s}}_k^T \bar{\mathbf{C}}_k^{-1} \bar{\mathbf{s}}_k\right), \quad (13)$$

and using

$$\mathbf{B}^{(m)} \mathbf{X}^{(m)} = \mathbf{S}^{(m)} \Leftrightarrow \mathbf{e}_k^T \mathbf{B}^{(m)} \mathbf{X}^{(m)} = \mathbf{s}_k^{(m)T}, \\ \forall k \in \{1, \dots, K\}, \forall m \in \{1, \dots, M\}, \quad (14)$$

we obtain the normalized log-likelihood of $\bar{\mathbf{B}}$, given by

$$\mathcal{L}(\bar{\mathbf{B}}) = \mathcal{L}(\mathbf{B}^{(1)}, \dots, \mathbf{B}^{(M)}) \triangleq \frac{1}{T} \log p_{\bar{\mathbf{X}}}(\bar{\mathbf{X}}; \bar{\mathbf{B}}) \\ = \gamma + \log |\det \bar{\mathbf{B}}| \\ - \frac{1}{2T} \sum_{k=1}^K \sum_{m_1=1}^M \sum_{m_2=1}^M \mathbf{e}_k^T \mathbf{B}^{(m_1)} \mathbf{X}^{(m_1)} \mathbf{P}_k^{(m_1, m_2)} \mathbf{X}^{(m_2)T} \mathbf{B}^{(m_2)T} \mathbf{e}_k \\ = \gamma + \sum_{\ell=1}^M \log |\det \mathbf{B}^{(\ell)}| \\ - \frac{1}{2} \sum_{k=1}^K \sum_{m_1=1}^M \sum_{m_2=1}^M \mathbf{e}_k^T \mathbf{B}^{(m_1)} \mathbf{Q}_k^{(m_1, m_2)} \mathbf{B}^{(m_2)T} \mathbf{e}_k, \quad (15)$$

where the matrices $\mathbf{Q}_k^{(m_1, m_2)}$ (to be later referred to as the target-matrices), are defined as

$$\mathbf{Q}_k^{(m_1, m_2)} \triangleq \frac{1}{T} \mathbf{X}^{(m_1)} \mathbf{P}_k^{(m_1, m_2)} \mathbf{X}^{(m_2)T}, \\ \forall k \in \{1, \dots, K\}, \forall m_1, m_2 \in \{1, \dots, M\}, \quad (16)$$

and where γ is a constant independent of $\bar{\mathbf{B}}$. Since the ML estimate $\hat{\mathbf{B}}_{\text{ML}}^{(1)}, \dots, \hat{\mathbf{B}}_{\text{ML}}^{(M)}$ of $\mathbf{B}^{(1)}, \dots, \mathbf{B}^{(M)}$ are the global maximizers of the likelihood function, we seek the solution of

$$\frac{\partial \mathcal{L}(\bar{\mathbf{B}})}{\partial \mathbf{B}^{(m)}} \stackrel{!}{=} \mathbf{O}, \quad \forall m \in \{1, \dots, M\}, \quad (17)$$

(where $\stackrel{!}{=}$ denotes a demand for equality) which corresponds to the global maximum, where $\mathbf{O} \in \mathbb{R}^{K \times K}$ is the all-zeros matrix. Indeed, differentiating $\mathcal{L}(\bar{\mathbf{B}})$ w.r.t. $\mathbf{B}^{(m)}$ and equating to zero yields (see Appendix A for details)

$$\frac{\partial \mathcal{L}(\bar{\mathbf{B}})}{\partial \mathbf{B}^{(m)}} \bigg|_{\hat{\bar{\mathbf{B}}}_{\text{ML}}} = \frac{\partial}{\partial \mathbf{B}^{(m)}} \left(\sum_{\ell=1}^M \log |\det \mathbf{B}^{(\ell)}| \right. \\ \left. - \frac{1}{2} \sum_{k=1}^K \sum_{m_1=1}^M \sum_{m_2=1}^M \mathbf{e}_k^T \mathbf{B}^{(m_1)} \mathbf{Q}_k^{(m_1, m_2)} \mathbf{B}^{(m_2)T} \mathbf{e}_k + \gamma \right) \bigg|_{\hat{\bar{\mathbf{B}}}_{\text{ML}}} \\ = \hat{\mathbf{A}}_{\text{ML}}^{(m)T} - \sum_{k=1}^K \sum_{\ell=1}^M \mathbf{E}_{kk} \hat{\mathbf{B}}_{\text{ML}}^{(\ell)} \mathbf{Q}_k^{(\ell, m)} \stackrel{!}{=} \mathbf{O}, \quad \forall m \in \{1, \dots, M\}, \quad (18)$$

where $\{\hat{\mathbf{A}}_{\text{ML}}^{(m)} \triangleq \hat{\mathbf{B}}_{\text{ML}}^{(m)-1}\}_{m=1}^M$ and $\mathbf{E}_{ij} \triangleq \mathbf{e}_i \mathbf{e}_j^T$. So

$$\sum_{k=1}^K \sum_{\ell=1}^M \mathbf{E}_{kk} \hat{\mathbf{B}}_{\text{ML}}^{(\ell)} \mathbf{Q}_k^{(\ell, m)} = \hat{\mathbf{A}}_{\text{ML}}^{(m)T}, \quad \forall m \in \{1, \dots, M\}. \quad (19)$$

Transposing and left-multiplying by $\widehat{\mathbf{B}}_{\text{ML}}^{(m)}$ we get

$$\sum_{k=1}^K \sum_{\ell=1}^M \widehat{\mathbf{B}}_{\text{ML}}^{(m)} \mathbf{Q}_k^{(m,\ell)} \widehat{\mathbf{B}}_{\text{ML}}^{(\ell)\text{T}} \mathbf{E}_{kk} = \mathbf{I}_K, \quad \forall m \in \{1, \dots, M\}$$

$$\Leftrightarrow$$

$$\sum_{\ell=1}^M \widehat{\mathbf{B}}_{\text{ML}}^{(m)} \mathbf{Q}_k^{(m,\ell)} \widehat{\mathbf{B}}_{\text{ML}}^{(\ell)\text{T}} \mathbf{e}_k \triangleq \mathbf{D}_k^{(m)} \mathbf{e}_k = \mathbf{e}_k,$$

$$\forall m \in \{1, \dots, M\}, \forall k \in \{1, \dots, K\}, \quad (20)$$

where we have used $\mathbf{Q}_k^{(m_1, m_2)} = \mathbf{Q}_k^{(m_2, m_1)\text{T}}$. The set of equations (20) (and, equivalently, (19)) constitutes the *extended SeDJoCo* problem. It is easy to see that for the particular case where $M = 1$, the problem boils down to the standard (single dataset) SeDJoCo problem [24]. We emphasize that like in standard SeDJoCo, this means that the k -th column of the matrix $\mathbf{D}_k^{(m)}$ should be “drilled” (namely be all-zeros except for its k -th element equaling 1), but unlike standard SeDJoCo, here the matrix $\mathbf{D}_k^{(m)}$ is not necessarily symmetric, in general, so the extended SeDJoCo transformation is not symmetric in the sense that the rows of $\mathbf{D}_k^{(m)}$ are generally not “drilled”. In other words, \mathbf{e}_k is a right-eigenvector of $\mathbf{D}_k^{(m)}$ (corresponding to the eigenvalue 1), whereas \mathbf{e}_k^{T} is not necessarily a left-eigenvector of $\mathbf{D}_k^{(m)}$.

When the source signals in different datasets are statistically independent, the matrices $\widehat{\mathbf{C}}_k$, as well as their inverses, become block-diagonal, so that all $\mathbf{P}_k^{(m_1, m_2)}$ (and therefore also all $\mathbf{Q}_k^{(m_1, m_2)}$) vanish for all $m_1 \neq m_2$. As a result, the problem reduces to a set of M “standard” (decoupled) SeDJoCo problems, since only one element (corresponding to $\ell = m$) is left in each sum in (20). However, in a “true” IVA setup sources from different datasets *are* correlated, giving rise to a non-degenerate extended SeDJoCo problem.

C. Justification of the Semi-Blindness Assumption

As already mentioned, our semi-blind scenario, which assumes prior knowledge of the full SCV covariance matrices (in addition to the Gaussianity assumption), may seem questionable. There are, however, several possible arguments in support of considering such a scenario - including a specific practical example.

- From a theoretical point of view: ML estimation always assumes prior knowledge of the full statistical model of the observations, up to the unknown parameters (to be estimated). Such knowledge is not always realistic in practice, but still ML estimation enjoys tremendous popularity as a baseline theoretical approach. Thus, one of our objectives is to derive the ML estimation in the IVA context with a general Gaussian model, and to show how prior knowledge of the SCVs’ covariances (which is sufficient in the Gaussian case) can be exploited in an optimal manner if and when it is available.
- Assuming a “training” period: In some applications the user might have access to the unmixed source signals during some “training” period, before actually observing the mixtures in the “operational” period. Assuming that

the statistical properties of the sources during the training period remain valid during the operational period, these properties may be estimated during the training period, to be used in turn during the operational period.

- A possible iterative scheme: When the covariance matrices are not known *a-priori*, but can be succinctly parameterized (e.g., in the case of stationary parametric auto-regressive / moving average sources), an iterative separation strategy may be used as follows: The sources are first estimated by some initial (non-ML) separation (if possible), which may not be optimal, but would still be reasonable enough to allow subsequent estimation of the required parameters for the covariance matrices from the separated signals. Then a semi-blind framework would be applied using the estimated covariance matrices, possibly with successive refinements by repeating the process - thereby approaching asymptotic optimality. In fact, it is also conceivable to operate such a process in an adaptive (rather than a batch) scheme, in which the estimation of the covariance matrices is interlaced with the estimation of the unmixing matrices as new measurements keep flowing in. However, this is a topic for further research, beyond the scope of the current paper.
 - A “real world” practical example: Consider K different sources (e.g., transmitters), remotely positioned at known locations, transmitting different, independent signals, which all have the same (known) spectrum and can be assumed Gaussian. This is a very common assumption, e.g., when the transmitters transmit different communication signals with the same standard modulation, which has a known spectrum (implying a known autocorrelation function). For example, Orthogonal Frequency Division Multiplexing (OFDM) digital communication signals are commonly modeled as Gaussian, see, e.g., [30]. Assume first that these signals are received at a single site by K different sensors (antennas) with different (unknown, or uncalibrated) radiation patterns, directed at different directions (not necessarily in the form of a calibrated phased array). The signal received at each antenna is then a different (unknown) linear mixture of all K sources. Note that the mixture is non-separable (without further information on the signals) since it consists of Gaussian sources with identical spectra.
- Now assume a second, distant reception site with K similar antennas. The source signals received at this site are differently-delayed versions of those from the other site, due to propagation delays. Knowing the positions of the sources and of the sensor sites, these delays can be calculated, and can be readily used to obtain the cross-correlations between respective sources at both sites, which is simply their autocorrelations, shifted by the respective (positive or negative) delay differences. Thus, the $2K$ signals at the two sites give rise to an IVA problem with $M = 2$ sets (easily extended to any M by adding more sites), in a semi-blind Gaussian scenario: all necessary covariance (and cross-covariance) matrices are known. Moreover, the signals at each site alone are non-separable - yet, using our semi-blind IVA scheme we can

obtain optimal separation of the sources.

The fine details of this scenario introduce some minor complications (e.g., in the context of communication signals, a complex-valued extension of our results should be used), so we shall not pursue this problem in here any further - however we did get good separation results with this scenario (even in the presence of additive noise), so this is at least one practical example where prior knowledge of the required covariance matrices is quite realistic.

Interestingly, we may also add that a very similar formulation to the extended SeDJoCo can again be linked to a CBF problem, but in an extended multi-cast framework, where M inter-connected transmitters, each with K antennas and K associated users are required to attain perfect interference cancellation transmission. This scenario is described in [31] in detail along with the mathematical derivation which leads to the following set of equations

$$\sum_{\ell=1}^M \mathbf{B}^{(\ell)} \mathbf{Q}_k^{(\ell,m)} \mathbf{B}^{(m)H} \mathbf{e}_k \triangleq \widetilde{\mathbf{D}}_k^{(m)} \mathbf{e}_k = \alpha \mathbf{e}_k, \quad \forall k \in \{1, \dots, K\} \quad \forall m \in \{1, \dots, M\}, \quad (21)$$

where in this case the target-matrices are defined as

$$\mathbf{Q}_k^{(m_1, m_2)} \triangleq \mathbf{H}_k^{(m_2, m_1)H} \mathbf{H}_k^{(m_2, m_2)} \in \mathbb{C}^{K \times K}, \quad \forall k \in \{1, \dots, K\} \quad \forall m_1, m_2 \in \{1, \dots, M\}, \quad (22)$$

where each matrix $\mathbf{H}_k^{(m_1, m_2)} \in \mathbb{C}^{K \times K}$ contains the flat fading channel coefficients from the K antennas of the m_2 -th transmitter to the K antennas of the k -th associated user of the m_1 -th transmitter. Evidently, these equations resemble the extended SeDJoCo equations (20) (with a change of the summation index; with a conjugate transpose replacing the transpose; and with the allowed scaling factor α). The solution of the resulting extended SeDJoCo problem would enable interference free delivery of the intended data streams to all users. Note that the only main difference of (21) from (20) is captured by the (different) definitions of the transformed matrices $\widetilde{\mathbf{D}}_k^{(m)}$ and $\mathbf{D}_k^{(m)}$.

These arguments serve as the basis for our motivation to delve deeper into this problem and to provide further results and insights regarding these equations.

The rest of this paper is structured as follows. In Section II we consider theoretical aspects of the problem like an alternative formulation, conditions for the existence of a solution and discussion of its non-uniqueness. In Section III we derive the induced Cramér-Rao lower bound (iCRLB, [21]) on the interference-to-signal ratio (ISR) for the underlying Gaussian IVA problem. In Section IV we propose two iterative solution algorithms for extended SeDJoCo. Comparative simulation results are presented in Section V, and Section VI is devoted to conclusions.

II. THEORETICAL ASPECTS OF EXTENDED SEDJOCo

A. Problem Formulations, Existence, Non-Uniqueness

We start with (re)formulating the (context-free) extended SeDJoCo problem. Recall that in this problem we con-

sider M sets, giving rise to KM^2 target-matrices, each of dimension $K \times K$, with a solution in the form of M $K \times K$ matrices. Hence, the extended SeDJoCo is stated as follows: Given KM^2 target-matrices $\{\mathbf{Q}_k^{(m_1, m_2)}\} \quad \forall k \in \{1, \dots, K\}, \forall m_1, m_2 \in \{1, \dots, M\}$,

P1: find a set of M $K \times K$ matrices $\{\mathbf{B}^{(m)}\}_{m=1}^M$, such that

$$\left[\sum_{\ell=1}^M \mathbf{B}^{(m)} \mathbf{Q}_k^{(m, \ell)} \mathbf{B}^{(\ell)T} \right] \mathbf{e}_k \triangleq \mathbf{D}_k^{(m)} \mathbf{e}_k = \mathbf{e}_k, \quad \forall k \in \{1, \dots, K\}, \forall m \in \{1, \dots, M\}. \quad (23)$$

The meaning of this statement is that the transformed matrices $\mathbf{D}_k^{(m)}$ should all be *exactly* “drilled” in their k -th column.

Equivalently, this problem can be stated as:

P2: find a set of KM vectors $\{\mathbf{b}_{k_1}^{(m)} \in \mathbb{R}^{K \times 1}\}, k \in \{1, \dots, K\}, m \in \{1, \dots, M\}$, such that

$$\sum_{\ell=1}^M \mathbf{b}_{k_1}^{(m)T} \mathbf{Q}_{k_2}^{(m, \ell)} \mathbf{b}_{k_2}^{(\ell)} = \delta_{k_1 k_2}, \quad \forall k_1, k_2 \in \{1, \dots, K\}, \forall m \in \{1, \dots, M\}, \quad (24)$$

where $\delta_{k_1 k_2}$ denotes Kronecker’s delta function (which is 1 if $k_1 = k_2$ and 0 otherwise).

This formulation suggests that each solution vector $\mathbf{b}_{k_1}^{(m)}$ (which is simply the k_1 -th row of $\mathbf{B}^{(m)}$) is orthogonal to all (transformed) vectors $\boldsymbol{\psi}_{k_2}^{(m)} \triangleq \sum_{\ell=1}^M \mathbf{Q}_{k_2}^{(m, \ell)} \mathbf{b}_{k_2}^{(\ell)}$ where $k_1 \neq k_2$ (for any $m \in \{1, \dots, M\}$). With these notations, we have that $\mathbf{B}^{(m)T} = [\mathbf{b}_1^{(m)} \dots \mathbf{b}_K^{(m)}]$ for all $m \in \{1, \dots, M\}$.

As seen from both formulations above, the extended SeDJoCo requires the solution of MK^2 equations in MK^2 unknowns, the elements of the matrices $\{\mathbf{B}^{(m)}\}_{m=1}^M$. Nevertheless, since these equations are nonlinear (in particular, they contain only 2nd order monomials of the unknowns), conclusions regarding the existence and/or uniqueness of the solution are non-trivial. The following is a sufficient condition for the existence of a (generally non-unique) solution. In the sequel we shall characterize a set of solutions which may exist when at least one solution exists.

Theorem 1 (a sufficient condition for existence of a solution): For a given set of target-matrices $\{\mathbf{Q}_k^{(m_1, m_2)}\} \in \mathbb{R}^{K \times K} \quad \forall k \in \{1, \dots, K\}, \forall m_1, m_2 \in \{1, \dots, M\}$, a solution for the associated extended SeDJoCo problem exists if all K matrices $\{\boldsymbol{\Omega}_k\}_{k=1}^K$, defined as

$$\boldsymbol{\Omega}_k \triangleq \begin{bmatrix} \mathbf{Q}_k^{(1,1)} & \dots & \mathbf{Q}_k^{(1,M)} \\ \vdots & \ddots & \vdots \\ \mathbf{Q}_k^{(M,1)} & \dots & \mathbf{Q}_k^{(M,M)} \end{bmatrix} \in \mathbb{R}^{KM \times KM}, \quad (25)$$

are Positive Definite (PD).

Proof 1: Let $\boldsymbol{\Omega}_1, \dots, \boldsymbol{\Omega}_K$ denote a set of (symmetric, real-valued) PD matrices, constructed from the target-matrices as defined in (25), and let $\lambda_k > 0$ denote the smallest eigenvalue of $\boldsymbol{\Omega}_k$, $k = 1, \dots, K$. Define

$$\widetilde{\mathbf{B}} \triangleq [\mathbf{B}^{(1)} \dots \mathbf{B}^{(M)}] \in \mathbb{R}^{K \times KM}, \quad (26)$$

and denote its rows as $\tilde{\mathbf{b}}_k^T, k = 1, \dots, K$. Now consider the function

$$C(\tilde{\mathbf{B}}) \triangleq \sum_{m=1}^M \log |\det \mathbf{B}^{(m)}| - \frac{1}{2} \sum_{k=1}^K \tilde{\mathbf{b}}_k^T \Omega_k \tilde{\mathbf{b}}_k. \quad (27)$$

For all nonsingular $\mathbf{B}^{(1)}, \dots, \mathbf{B}^{(M)}$, $C(\tilde{\mathbf{B}})$ is obviously a continuous and differentiable function of all elements of all the matrices. In addition, $C(\tilde{\mathbf{B}})$ is bounded from above:

$$\begin{aligned} C(\tilde{\mathbf{B}}) &= \sum_{m=1}^M \log |\det \mathbf{B}^{(m)}| - \frac{1}{2} \sum_{k=1}^K \tilde{\mathbf{b}}_k^T \Omega_k \tilde{\mathbf{b}}_k \\ &\leq \sum_{m=1}^M \log \prod_{k=1}^K \|\mathbf{b}_k^{(m)}\| - \frac{1}{2} \sum_{k=1}^K \lambda_k \tilde{\mathbf{b}}_k^T \tilde{\mathbf{b}}_k \\ &= \frac{1}{2} \sum_{m=1}^M \sum_{k=1}^K \{\log \|\mathbf{b}_k^{(m)}\|^2 - \lambda_k \|\mathbf{b}_k^{(m)}\|^2\} \\ &\leq \frac{1}{2} \sum_{m=1}^M \sum_{k=1}^K \{-\log \lambda_k - 1\} \\ &= \frac{M}{2} \sum_{k=1}^K \{\log \lambda_k - 1\}, \end{aligned} \quad (28)$$

where $\|\cdot\|$ denotes the ℓ^2 -norm, and where we have used the properties

- 1) $|\det \mathbf{B}^{(m)}| \leq \prod_{k=1}^K \|\mathbf{b}_k^{(m)}\|$ (Hadamard's inequality);
- 2) $\tilde{\mathbf{b}}_k^T \Omega_k \tilde{\mathbf{b}}_k \geq \lambda_k \|\tilde{\mathbf{b}}_k\|^2$;
- 3) $\|\tilde{\mathbf{b}}_k\|^2 = \sum_{m=1}^M \|\mathbf{b}_k^{(m)}\|^2$; and
- 4) $\log x - \lambda x \leq -\log \lambda - 1$ for all $x > 0$.

Note also that $C(\tilde{\mathbf{B}})$ tends to $-\infty$ when at least one of the matrices $\{\mathbf{B}^{(m)}\}_{m=1}^M$ approaches any singular matrix, and that, in addition, $C(\tilde{\mathbf{B}})$ has the property

$$C(\alpha \cdot \tilde{\mathbf{B}}) \xrightarrow{\alpha \rightarrow \infty} -\infty, \quad \forall \tilde{\mathbf{B}} \in \mathbb{R}^{K \times KM}. \quad (29)$$

Consequently, $C(\tilde{\mathbf{B}})$ must attain a maximum for (at least) some set of nonsingular matrices $\{\mathbf{B}^{(m)}\}_{m=1}^M$. Being a smooth function thereof derivative w.r.t. each $\{\mathbf{B}^{(m)}\}_{m=1}^M$ at the maximum point must vanish.

Indeed, differentiating $C(\tilde{\mathbf{B}})$ w.r.t. $\mathbf{B}^{(m)}$ we get

$$\begin{aligned} \frac{\partial C(\tilde{\mathbf{B}})}{\partial \mathbf{B}^{(m)}} &= \mathbf{A}^{(m)T} - \frac{1}{2} \frac{\partial}{\partial \mathbf{B}^{(m)}} \left[\sum_{k=1}^K \mathbf{e}_k^T \tilde{\mathbf{B}} \Omega_k \tilde{\mathbf{B}}^T \mathbf{e}_k \right] \\ &= \mathbf{A}^{(m)T} - \frac{1}{2} \frac{\partial}{\partial \mathbf{B}^{(m)}} \left[\sum_{k=1}^K \sum_{\ell=1}^M \sum_{p=1}^M \mathbf{e}_k^T \mathbf{B}^{(\ell)} \mathbf{Q}_k^{(\ell,p)} \mathbf{B}^{(p)T} \mathbf{e}_k \right] \\ &= \mathbf{A}^{(m)T} - \frac{1}{2} \sum_{k=1}^K \sum_{\ell=1}^M 2 \mathbf{E}_{kk} \mathbf{B}^{(\ell)} \mathbf{Q}_k^{(\ell,m)} \\ &= \mathbf{A}^{(m)T} - \sum_{k=1}^K \sum_{\ell=1}^M \mathbf{E}_{kk} \mathbf{B}^{(\ell)} \mathbf{Q}_k^{(\ell,m)}, \quad \forall m \in \{1, \dots, M\}, \end{aligned} \quad (30)$$

where $\mathbf{A}^{(m)} \triangleq \mathbf{B}^{(m)-1}$, and we have used the properties stated in Appendix A. And so, equating to zero, transposing and left-multiplying by $\mathbf{B}^{(m)}$, we get

$$\sum_{k=1}^K \sum_{\ell=1}^M \mathbf{B}^{(m)} \mathbf{Q}_k^{(m,\ell)} \mathbf{B}^{(\ell)T} \mathbf{E}_{kk} = \mathbf{I}_K, \quad \forall m \in \{1, \dots, M\} \quad (31)$$

which implies (23). This means that a solution of extended SeDJoCo must exist as the maximizer of $C(\tilde{\mathbf{B}})$, as long as $\{\Omega_k\}_{k=1}^K$ are all PD. \square

Note that this general result holds for any extended SeDJoCo problem, and is not limited to the context of IVA. For an IVA problem, when the $\mathbf{Q}_k^{(m_1, m_2)}$ matrices are given by (16), it is easy to show that if all the SCVs covariance matrices $\tilde{\mathbf{C}}_k$ are PD and finite, so are all Ω_k , with probability 1 (w.p.1). To observe this, note that in this case each Ω_k can be expressed as $\Omega_k = \mathcal{X} \bar{\mathbf{P}}_k \mathcal{X}^T$, where $\mathcal{X} \triangleq \text{Bdiag}(\mathbf{X}^{(1)}, \dots, \mathbf{X}^{(M)}) \in \mathbb{R}^{KM \times MT}$. If $\tilde{\mathbf{C}}_k$ is PD and finite, so is $\bar{\mathbf{P}}_k$, and since \mathcal{X} is full rank w.p.1, Ω_k is also PD (w.p.1). It is important to realize, however, that existence of an extended SeDJoCo solution in a given IVA problem does not necessarily imply separability, because an infinite number of solutions may exist over a manifold in the parameters space. In fact, this is what happens when the identifiability conditions stated in [6] are not satisfied, e.g., when two Gaussian SCVs have the same covariance matrices in which case the iCRLB is infinite for the associated ISRs. When the identifiability conditions in [6] are satisfied, then although the SeDJoCo solution is still not unique (see below), the multiple solutions are all isolated (w.p.1), and only one corresponds to the global maximum of the likelihood function.

Having addressed the issue of existence of a solution, we proceed to discuss the issue of uniqueness. Note first, that since, as mentioned above, extended SeDJoCo is a system of MK^2 equations in MK^2 unknowns (elements of $\tilde{\mathbf{B}}$), where each equation is a second degree multinomial in the unknowns, Bézout's theorem (e.g., [32]) asserts that there are at most 2^{MK^2} distinct real-valued solutions. Indeed, according to our experience, when a solution exists, it is not unique, in general. Moreover, we can generally characterize $K!$ essentially different solutions whenever a single solution exists.

To this end, assume an extended SeDJoCo problem associated with a set of target-matrices $\{\mathbf{Q}_k^{(m_1, m_2)}\}$ that satisfy the existence condition (all implied Ω_k matrices are PD). Denote by $\{\mathbf{B}^{(m)}\}_{m=1}^M$ a solution to this problem. Now define a new set of target-matrices $\{\bar{\mathbf{Q}}_k^{(m_1, m_2)}\}$, such that $(\forall m_1, m_2 \in \{1, \dots, M\})$

$$\bar{\mathbf{Q}}_1^{(m_1, m_2)} = \mathbf{Q}_2^{(m_1, m_2)}, \quad \bar{\mathbf{Q}}_2^{(m_1, m_2)} = \mathbf{Q}_1^{(m_1, m_2)}, \quad (32)$$

and

$$\bar{\mathbf{Q}}_k^{(m_1, m_2)} = \mathbf{Q}_k^{(m_1, m_2)} \quad \forall k \in \{3, \dots, K\}, \quad (33)$$

thereby defining a modified (permuted) extended SeDJoCo problem. Obviously, the modified problem also has at least one solution (the existence condition is still satisfied), but a solution of the original problem generally does not solve this

modified problem. However, an iterative algorithm starting at the solution $\{\mathbf{B}^{(m)}\}_{m=1}^M$ of the original problem is likely to reach an essentially different solution of the modified problem, in which all elements of the resulting solution matrices are generally different from all elements of the solution matrices of the original problem, because different coefficients now multiply different products of the unknowns in the system of equations. Thus, let us denote by $\{\overline{\mathbf{B}}^{(m)}\}_{m=1}^M$ the resulting solution of the modified problem.

Now let $\Pi_{1,2} \in \mathbb{R}^{K \times K}$ denote the (symmetric) permutation matrix that swaps the first and second elements of a vector, namely $\Pi_{1,2}\mathbf{e}_1 = \mathbf{e}_2$, $\Pi_{1,2}\mathbf{e}_2 = \mathbf{e}_1$ and $\Pi_{1,2}\mathbf{e}_k = \mathbf{e}_k$ for all other $k \in \{3, \dots, K\}$. Consider the set of matrices $\{\mathbf{B}'^{(m)} \triangleq \Pi_{1,2}\overline{\mathbf{B}}^{(m)}\}$ for all $m \in \{1, \dots, M\}$. We assert that this set of matrices solves the extended SeDJoCo problem induced by the *original* set $\{\mathbf{Q}_k^{(m_1, m_2)}\}$, since

$$\begin{aligned} & \left[\sum_{m=1}^M \mathbf{B}'^{(\ell)} \mathbf{Q}_1^{(\ell, m)} \mathbf{B}'^{(m)T} \right] \mathbf{e}_1 \\ &= \Pi_{1,2} \left[\sum_{m=1}^M \overline{\mathbf{B}}^{(\ell)} \overline{\mathbf{Q}}_2^{(\ell, m)} \overline{\mathbf{B}}^{(m)T} \right] \Pi_{1,2}^T \mathbf{e}_1 \\ &= \Pi_{1,2} \left[\sum_{m=1}^M \overline{\mathbf{B}}^{(\ell)} \overline{\mathbf{Q}}_2^{(\ell, m)} \overline{\mathbf{B}}^{(m)T} \right] \mathbf{e}_2 \\ &= \Pi_{1,2} \mathbf{e}_2 = \mathbf{e}_1, \quad \forall \ell \in \{1, \dots, M\}, \end{aligned} \quad (34)$$

and, similarly, $\left[\sum_{m=1}^M \mathbf{B}'^{(\ell)} \mathbf{Q}_2^{(\ell, m)} \mathbf{B}'^{(m)T} \right] \mathbf{e}_2 = \mathbf{e}_2$, and, of course, $\left[\sum_{m=1}^M \mathbf{B}'^{(\ell)} \mathbf{Q}_k^{(\ell, m)} \mathbf{B}'^{(m)T} \right] \mathbf{e}_k = \mathbf{e}_k$ for all other $k \in \{3, \dots, K\}$ (for all $\ell \in \{1, \dots, M\}$). This means that in addition to $\{\mathbf{B}^{(m)}\}_{m=1}^M$ there exists an additional solution $\{\mathbf{B}'^{(m)}\}_{m=1}^M$ to the original extended SeDJoCo problem, which is a permutation of an “essentially different” solution $\{\overline{\mathbf{B}}^{(m)}\}_{m=1}^M$ to a permuted extended SeDJoCo problem. Since any permutation matrix can be expressed as the product of two-elements-permutation matrices, we may generalize the above result, i.e., there may exist, in general, $K!$ (the number of possible permutations) such different solutions for a K -dimensional extended SeDJoCo problem.

Note that, strictly speaking, we did not *prove* that the permuted extended SeDJoCo problem yields an *essentially different* solution, since theoretically the resulting solution of the permuted problem may just be a permuted version of the solution to the original problem. However, based on our empirical experience, we conjecture that with randomly generated target matrices (such as in IVA), an iterative algorithm starting at a solution of the original problem would “almost surely” reach an essentially different (not just permuted) solution of the permuted problem. In the context of our IVA problem, all these $K!$ different solutions would be local maxima of the Likelihood function, but (w.p.1) only one of them would correspond to the global maximum, and may be found using strategies such as those advocated (in the context of ICA) in [33], [34].

III. ICRLB ON THE ISR FOR JBSS

The ISR is a common measure in BSS which quantifies the “quality” of separation. More specifically, by definition

$$\text{ISR}_{k\ell}^{(m)} \triangleq E \left[\frac{\left| \left(\widehat{\mathbf{B}}^{(m)} \mathbf{A}^{(m)} \right)_{k\ell} \right|^2}{\left| \left(\widehat{\mathbf{B}}^{(m)} \mathbf{A}^{(m)} \right)_{kk} \right|^2} \right] \cdot \frac{E \left[\mathbf{s}_\ell^{(m)T} \mathbf{s}_\ell^{(m)} \right]}{E \left[\mathbf{s}_k^{(m)T} \mathbf{s}_k^{(m)} \right]}, \quad \forall k, \ell \in \{1, \dots, K\}, k \neq \ell, \forall m \in \{1, \dots, M\}, \quad (35)$$

measures the expected relative residual energy of the ℓ -th source in the reconstruction of the k -th source in the m -th dataset.

By deriving the CRLB on the estimation of the mixing matrix and using the equivariance property, it is possible to obtain the iCRLB on the ISR (e.g., [6], [21], [23], [35]). In [6] general expressions for the iCRLB are provided in the context of a general IVA problem. It is shown that

$$\text{ISR}_{k\ell}^{(m)} \geq \frac{1}{T} \cdot \mathbf{e}_m^T [\mathcal{K}_{k\ell} - \mathcal{K}_{\ell k}^{-1}]^{-1} \mathbf{e}_m \cdot \frac{\text{Tr}(\mathbf{C}_\ell^{(m, m)})}{\text{Tr}(\mathbf{C}_k^{(m, m)})}, \quad (36)$$

where $\text{Tr}(\cdot)$ denotes the trace operator, and where the elements of the matrices $\mathcal{K}_{k\ell} \in \mathbb{R}^{M \times M}$ are defined as

$$\mathcal{K}_{k\ell}[m, n] \triangleq \frac{1}{T} \text{Tr}(\mathbf{\Gamma}_k^{(m, n)} \mathbf{C}_\ell^{(m, n)}), \quad (37)$$

in which the matrices $\mathbf{\Gamma}_k^{(m, n)} \in \mathbb{R}^{T \times T}$ are defined as

$$\mathbf{\Gamma}_k^{(m, n)} \triangleq E \left[\phi_k^{(m)T} \phi_k^{(n)} \right]. \quad (38)$$

Here $\phi_k^{(m)} \in \mathbb{R}^{1 \times T}$ denotes the score vector of the m -th component of the k -th source, namely the derivative of the (negative) log of the probability distribution of the k -th SCV $\bar{\mathbf{s}}_k$ w.r.t. its m -th component $\mathbf{s}_k^{(m)}$. Different IVA models naturally have different score vectors, giving rise to different $\mathbf{\Gamma}_k^{(m, n)}$ matrices and, thereby, to different iCRLBs. In order to obtain an explicit expression for the iCRLB in a given IVA model, the respective $\mathbf{\Gamma}_k^{(m, n)}$ matrices need to be explicitly calculated. In [6] the authors derive explicit results for the simple case of temporally independent, identically distributed (i.i.d.) Gaussian sources. To obtain $\mathbf{\Gamma}_k^{(m, n)}$ for our more general temporal models, recall that

$$\begin{aligned} -\log p_{\bar{\mathbf{s}}_k}(\bar{\mathbf{s}}_k) &= \frac{1}{2} \log \det |2\pi \bar{\mathbf{C}}_k| + \frac{1}{2} \bar{\mathbf{s}}_k^T \bar{\mathbf{C}}_k^{-1} \bar{\mathbf{s}}_k \\ &= \frac{1}{2} \log \det |2\pi \bar{\mathbf{C}}_k| + \frac{1}{2} \bar{\mathbf{s}}_k^T \bar{\mathbf{P}}_k \bar{\mathbf{s}}_k, \end{aligned} \quad (39)$$

so that its derivative w.r.t. the entire SCV $\bar{\mathbf{s}}_k$ is given by $\bar{\mathbf{s}}_k^T \bar{\mathbf{P}}_k \in \mathbb{R}^{1 \times MT}$. Its m -th component is therefore given by

$$\phi_k^{(m)} = \sum_{p=1}^M \mathbf{s}_k^{(p)T} \mathbf{P}_k^{(p, m)}. \quad (40)$$

Substituting in (38) we get

$$\begin{aligned}\Gamma_k^{(m,n)} &= E \left[\phi_k^{(m)T} \phi_k^{(n)} \right] = E \left[\sum_{p,q=1}^M \mathbf{P}_k^{(m,p)} \mathbf{s}_k^{(p)} \mathbf{s}_k^{(q)T} \mathbf{P}_k^{(q,n)} \right] \\ &= \sum_{p,q=1}^M \mathbf{P}_k^{(m,p)} \mathbf{C}_k^{(p,q)} \mathbf{P}_k^{(q,n)} = \sum_{q=1}^M \mathbf{I}_{MT}^{(m,q)} \mathbf{P}_k^{(q,n)} = \mathbf{P}_k^{(m,n)},\end{aligned}\quad (41)$$

where \mathbf{I}_{MT} denotes the $MT \times MT$ identity matrix and $\mathbf{I}_{MT}^{(m,q)}$ denotes its (m, q) -th $T \times T$ block (which is \mathbf{I}_T for $q = m$ and all zeros otherwise). To conclude, we can now substitute this result into (37), obtaining

$$\mathcal{K}_{k\ell}[m, n] = \frac{1}{T} \text{Tr} \left(\mathbf{P}_k^{(n,m)} \mathbf{C}_\ell^{(m,n)} \right), \quad (42)$$

which can in turn be substituted into (36) to yield the iCRLB.

IV. SOLVING EXTENDED SEDJoCo

To the best of our knowledge, the extended SeDJoCo problem, and in particular its solution, have not yet been addressed in the literature (excluding our recent conference papers [31], [36], in which the extended SeDJoCo was first formulated, and a partial solution, which ignores some of the target matrices, was proposed). In what follows we propose two comprehensive, general solution approaches, both based on extensions of existing iterative solutions of standard SeDJoCo. The first is an extension of the Iterative Relaxation (IR) proposed by Dégerine and Zaïdi [37]. The second is based on Newton's method.

Note that both algorithms rely on some initial guess. A plausible option for obtaining an initial guess would be to initialize each $\mathbf{B}^{(m)}$ to the solution of the respective standard SeDJoCo problem associated with the m -th set (thereby ignoring the information in the inter-set dependence).

A. Solution by Iterative Relaxations

Recall the second formulation of the extended SeDJoCo problem (24), and notice that it can be written as

$$\begin{aligned}\sum_{\ell=1}^M \mathbf{b}_{k_1}^{(m)T} \mathbf{Q}_{k_2}^{(m,\ell)} \mathbf{b}_{k_2}^{(\ell)} &= \mathbf{b}_{k_1}^{(m)T} \sum_{\ell=1}^M \mathbf{Q}_{k_2}^{(m,\ell)} \mathbf{b}_{k_2}^{(\ell)} \\ &= \mathbf{b}_{k_1}^{(m)T} \boldsymbol{\psi}_{k_2}^{(m)} = \delta_{k_1 k_2}, \\ \forall k_1, k_2 &\in \{1, \dots, K\}, \forall m \in \{1, \dots, M\}.\end{aligned}\quad (43)$$

This means that the vector $\mathbf{b}_k^{(m)}$ is orthogonal to all vectors $\{\boldsymbol{\psi}_{k'}^{(m)}\}_{k' \neq k}$ and its inner product with $\boldsymbol{\psi}_k^{(m)}$ equals

1. Therefore, if we assume that all vectors $\{\mathbf{b}_{k'}^{(m)}\}_{k' \neq k}$ are fixed, we can update the vector $\mathbf{b}_k^{(m)}$ by a somewhat-similar Gram-Schmidt procedure of subtracting its projection on the $K - 1$ subspace spanned by these vectors, followed by a "normalization" of the remaining residual. More precisely, the updating rule is as follows:

$$\begin{aligned}\mathbf{b}_k^{(m)'} &\leftarrow \mathbf{b}_k^{(m)} - \boldsymbol{\Psi}_{(k)}^{(m)T} \left(\boldsymbol{\Psi}_{(k)}^{(m)} \boldsymbol{\Psi}_{(k)}^{(m)T} \right)^{-1} \boldsymbol{\Psi}_{(k)}^{(m)} \mathbf{b}_k^{(m)} \\ \mathbf{b}_k^{(m)} &\leftarrow \mathbf{b}_k^{(m)'} / \sqrt{|\mathbf{b}_k^{(m)'}|^T \mathbf{b}_k^{(m)'}}\end{aligned}\quad (44)$$

where the rows of the matrix $\boldsymbol{\Psi}_{(k)}^{(m)} \in \mathbb{R}^{K-1 \times K}$ are the $K - 1$ vectors $\{\boldsymbol{\psi}_{k'}^{(m)}\}_{k' \neq k}$. The update rule is repeated iteratively until convergence, running through all $m \in \{1, \dots, M\}$ and $k \in \{1, \dots, K\}$.

B. Solution by Newton's Method

Let us define the gradient matrix of $\mathcal{L}(\bar{\mathbf{B}})$ as \mathbf{G} , such that

$$\begin{aligned}\frac{\partial \mathcal{L}(\bar{\mathbf{B}})}{\partial \bar{\mathbf{B}}} &\triangleq \mathbf{G} = \begin{bmatrix} \frac{\partial \mathcal{L}(\bar{\mathbf{B}})}{\partial \mathbf{B}^{(1)}} & \dots & \frac{\partial \mathcal{L}(\bar{\mathbf{B}})}{\partial \mathbf{B}^{(M)}} \end{bmatrix} \\ &\triangleq [\mathbf{G}^{(1)} \dots \mathbf{G}^{(M)}] \in \mathbb{R}^{K \times KM},\end{aligned}\quad (45)$$

such that $\mathbf{G}^{(\ell)} \in \mathbb{R}^{K \times K}, \forall \ell \in \{1, \dots, M\}$. Differentiating further w.r.t. $\mathbf{B}^{(n)}$, we get the block-matrices of the Hessian

$$\mathbf{H}^{(n,\ell)} \triangleq \frac{\partial \mathcal{L}(\bar{\mathbf{B}})}{\partial \mathbf{B}^{(n)} \partial \mathbf{B}^{(\ell)}} \in \mathbb{R}^{K^2 \times K^2}, \quad \forall n, \ell \in \{1, \dots, M\}, \quad (46)$$

where

$$\mathbf{H} \triangleq \begin{bmatrix} \mathbf{H}^{(1,1)} & \dots & \mathbf{H}^{(1,M)} \\ \vdots & \ddots & \vdots \\ \mathbf{H}^{(M,1)} & \dots & \mathbf{H}^{(M,M)} \end{bmatrix} \in \mathbb{R}^{K^2 M \times K^2 M}. \quad (47)$$

More explicitly, we have

$$\begin{aligned}\mathbf{H}^{(n,\ell)} &= \frac{\partial}{\partial \mathbf{B}^{(n)}} \mathbf{G}^{(\ell)} \\ &= \frac{\partial}{\partial \mathbf{B}^{(n)}} \left(\mathbf{A}^{(\ell)T} - \sum_{k=1}^K \sum_{m=1}^M \mathbf{E}_{kk} \mathbf{B}^{(m)} \mathbf{Q}_k^{(m,\ell)} \right) \\ &= - \left(\mathbf{I}_K \otimes \mathbf{A}^{(\ell)T} \right) \frac{\partial \mathbf{B}^{(\ell)T}}{\partial \mathbf{B}^{(n)}} \left(\mathbf{I}_K \otimes \mathbf{A}^{(\ell)T} \right) \\ &\quad - \sum_{k=1}^K \sum_{m=1}^M \left(\mathbf{I}_K \otimes \mathbf{E}_{kk} \right) \frac{\partial \mathbf{B}^{(m)}}{\partial \mathbf{B}^{(n)}} \left(\mathbf{I}_K \otimes \mathbf{Q}_k^{(m,\ell)} \right) \\ &= - \left(\mathbf{I}_K \otimes \mathbf{A}^{(\ell)T} \right) \left(\delta_{n\ell} \tilde{\mathbf{E}}^K \right) \left(\mathbf{I}_K \otimes \mathbf{A}^{(\ell)T} \right) \\ &\quad - \sum_{k=1}^K \sum_{m=1}^M \left(\mathbf{I}_K \otimes \mathbf{E}_{kk} \right) \left(\delta_{nm} \mathbf{E}^K \right) \left(\mathbf{I}_K \otimes \mathbf{Q}_k^{(m,\ell)} \right) \\ &= - \delta_{n\ell} \left(\mathbf{I}_K \otimes \mathbf{A}^{(\ell)T} \right) \tilde{\mathbf{E}}^K \left(\mathbf{I}_K \otimes \mathbf{A}^{(\ell)T} \right) \\ &\quad - \sum_{k=1}^K \left(\mathbf{I}_K \otimes \mathbf{E}_{kk} \right) \mathbf{E}^K \left(\mathbf{I}_K \otimes \mathbf{Q}_k^{(n,\ell)} \right),\end{aligned}\quad (48)$$

where we have used the property $\partial \mathbf{A} = -\mathbf{A} \partial \mathbf{B} \mathbf{A}$ [38], and the notations

$$\frac{\partial \mathbf{B}^{(\ell)}}{\partial \mathbf{B}^{(n)}} = \delta_{n\ell} \begin{bmatrix} \mathbf{E}_{11} & \dots & \mathbf{E}_{1M} \\ \vdots & \ddots & \vdots \\ \mathbf{E}_{M1} & \dots & \mathbf{E}_{MM} \end{bmatrix} \triangleq \delta_{n\ell} \mathbf{E}^K \in \mathbb{R}^{K^2 \times K^2}, \quad (49)$$

and

$$\frac{\partial \mathbf{B}^{(m)T}}{\partial \mathbf{B}^{(n)}} = \delta_{nm} \begin{bmatrix} \mathbf{E}_{11} & \dots & \mathbf{E}_{M1} \\ \vdots & \ddots & \vdots \\ \mathbf{E}_{1M} & \dots & \mathbf{E}_{MM} \end{bmatrix} \triangleq \delta_{nm} \tilde{\mathbf{E}}^K \in \mathbb{R}^{K^2 \times K^2}. \quad (50)$$

If we denote the columns of the matrix $\tilde{\mathbf{B}}$ by $\tilde{\beta}_j$ for $1 \leq j \leq KM$ and define the indexing function $j_{[k,m]} = (m-1) \cdot K + k$ for $1 \leq k \leq K$ and $1 \leq m \leq M$, we have that

$$\tilde{\beta}_{j_{[k,m]}} = \tilde{\beta}_k^{(m)}, \quad (51)$$

where $\tilde{\beta}_k^{(m)}$ denotes the k -th column of the matrix $\mathbf{B}^{(m)}$. Consequently, we conclude that the vectorized gradient of $\text{vec}(\tilde{\mathbf{B}})$ is $\text{vec}(\mathbf{G})$, for which \mathbf{G} is defined exactly as in (45). However, the Hessian matrix of the vector $\text{vec}(\tilde{\mathbf{B}})$ is given by $\tilde{\mathbf{H}}$, which is a permuted version of \mathbf{H} . More particularly, the elements of $\tilde{\mathbf{H}}$ are given by

$$\begin{aligned} \tilde{\mathbf{H}}_{(\text{ind}[p,q,i], \text{ind}[m,n,j])} &= \mathbf{H}_{(\text{ind}[p,m,i], \text{ind}[q,n,j])} \\ &= \frac{\partial \mathcal{L}(\tilde{\mathbf{B}})}{\partial \mathbf{B}_{(p,q)}^{(i)} \partial \mathbf{B}_{(m,n)}^{(j)}}, \end{aligned} \quad (52)$$

and the subscript (p,q) denotes the (p,q) -th element of a matrix. Finally, the elements of the matrices $\mathbf{B}^{(1)}, \dots, \mathbf{B}^{(M)}$ are computed iteratively according to Newton's update rule

$$\text{vec}(\tilde{\mathbf{B}})^{[n+1]} = \text{vec}(\tilde{\mathbf{B}})^{[n]} + \Delta^{[n]}, \quad (53)$$

where

$$\Delta^{[n]} = -\tilde{\mathbf{H}}^{-1} \cdot \text{vec}(\mathbf{G}) \in \mathbb{R}^{K^2 M} \quad (54)$$

evaluated at $\text{vec}(\tilde{\mathbf{B}})^{[n]}$, and the superscript $^{[n]}$ indicates the n -th iteration.

V. SIMULATION RESULTS

We present simulation results of 3 different experiments. First, we demonstrate the proposed algorithms' convergence behavior for a generic extended SeDJoCo problem with random target-matrices. We then proceed to demonstrate the performance of the extended SeDJoCo solution as the ML estimate in the context of Gaussian JBSS (IVA) in terms of common separation measures.

A. Convergence Behavior

In our first experiment we assess the convergence behavior of the two solutions proposed in Section IV for a generic extended SeDJoCo problem. In this experiment the target-matrices are generated as follows: First, we generate a set of K PD matrices by

$$\mathbf{\Omega}_k = \mathbf{U}_k \mathbf{U}_k^T \in \mathbb{R}^{KM \times KM}, \quad \forall k \in \{1, \dots, K\}, \quad (55)$$

where the elements of $\{\mathbf{U}_k\}_{k=1}^K$ are drawn independently from the standard Gaussian distribution. Then we take the $K \times K$ KM^2 blocks of all matrices $\{\mathbf{\Omega}_k\}_{k=1}^K$, as defined in (25), to be the set of target-matrices. This way a solution is guaranteed to exist (according to Theorem 1). We initialize the solution to be the set of identity matrices, i.e., $\{\mathbf{B}^{(m)} = \mathbf{I}_K\}_{m=1}^M$. Our measure of convergence is (cf. (31)) the logarithm of

$$\mathcal{E} \triangleq \left\| \sum_{m=1}^M \sum_{k=1}^K \sum_{\ell=1}^M \mathbf{B}^{(m)} \mathbf{Q}_k^{(m,\ell)} \mathbf{B}^{(\ell)T} \mathbf{E}_{kk} - \mathbf{I}_K \right\|_F, \quad (56)$$

which is the Frobenius norm of the residual-error matrix.

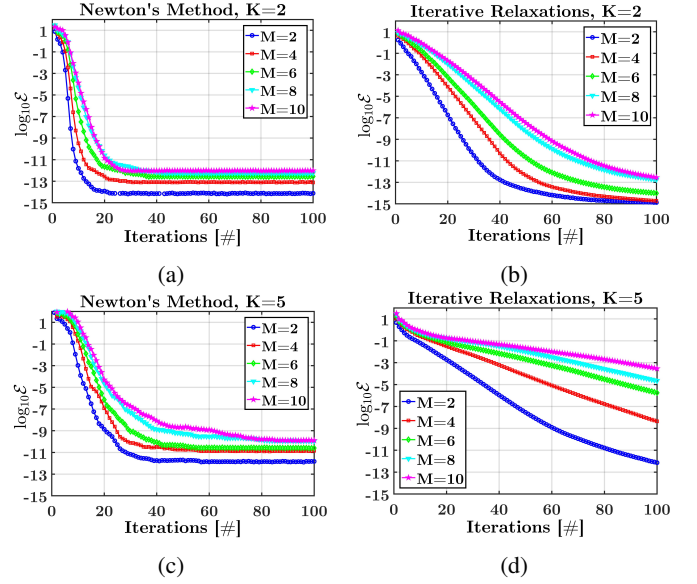


Fig. 2: Convergence patterns of the proposed algorithm - logarithm of the residual root-mean-squares error vs. the number of iterations for different values of M with a fixed value of K . (a) Newton's method, $K = 2$ (b) IR, $K = 2$ (c) Newton's method, $K = 5$ (d) IR, $K = 5$.

Fig. 2 shows the convergence patterns of the proposed algorithms for different values of M with K fixed. The results in Figs. 2a-2b and 2c-2d were obtained by averaging 100 independent identical trials for $K = 2$ and $K = 5$, respectively. It is evident that both algorithms converge to the solution; the Newton's algorithm converges much faster, typically within tens of iteration, whereas the IR algorithm converges significantly more slowly as K or M or both increase. For this example, with $K = 5$, $M = 4$, and when convergence is defined at $\mathcal{E} = 10^{-10}$, the average run time for convergence was 0.5457 seconds with Newton's method, and 1.7762 seconds with the IR algorithm, so that the solution by Newton's method was computationally more efficient in this case. Note, however, that the fast (quadratic) convergence of Newton's algorithm comes at the cost of a computational complexity increase per iteration; if we define a full update iteration as an update of all the elements of $\tilde{\mathbf{B}}$, the iterative relaxations algorithm requires $\mathcal{O}(MK^4)$ operations for a full update iteration (due to MK vector updates of $\mathcal{O}(K^3)$), whereas Newton's algorithm requires $\mathcal{O}(M^3 K^6)$ (due to inversion of the Hessian). In addition, and as expected, it can be seen that as K or M (or both) increase, the number of iterations increases as well (for both algorithms).

B. JBSS (IVA) of Gaussian Sources

In this part we focus on the application of the extended SeDJoCo as the ML solution for semi-blind Gaussian JBSS (where the sources' covariance matrices are assumed to be known in advance). Results are based on averaging 1000 independent trials, where the mixing matrices elements are redrawn from a standard Gaussian distribution in each trial. We solve the extended SeDJoCo problem with the solution by Newton's method due to its (empirical) faster convergence and better resilience to initialization in convergence to the ML

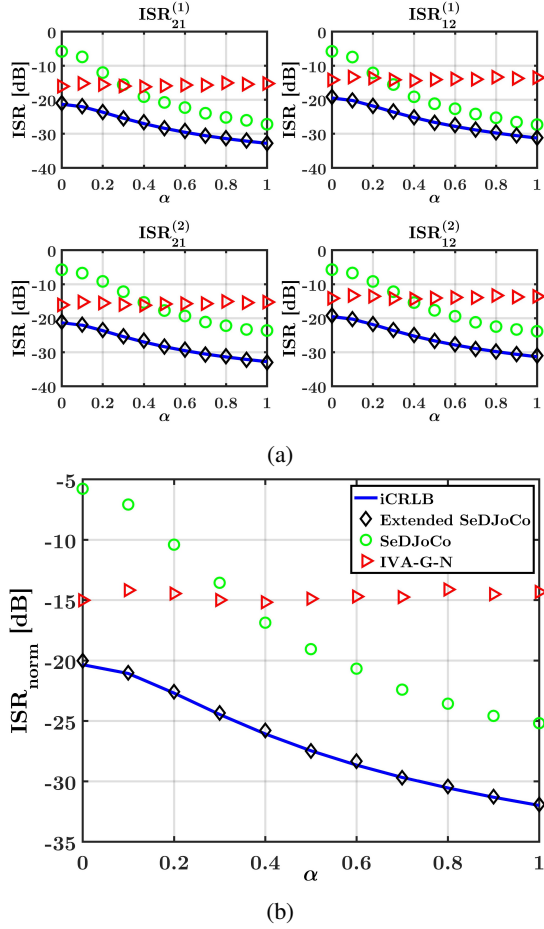


Fig. 3: (a) Empirical ISR vs. α (b) Empirical ISR_{norm} vs. α . It can be seen that the extended SeDJoCo solution is superior to both other solution. Furthermore, it attains the iCRLB. The legend in (b) is valid for (a) as well.

Parameter \ (k, m)	(1, 1)	(1, 2)	(2, 1)	(2, 2)
$\phi_0^{(k,m)}$	π	$\frac{5\pi}{3}$	$\frac{\pi}{3}$	π
$N_{k,m}$	50	350	200	500
$\sigma_k^{(m)}$	2	$3\frac{1}{3}$	$2\frac{2}{3}$	4

TABLE I: Fixed parameter values for experiment 2.

solution. First, we consider the simple case of two datasets ($M = 2$), each with two Gaussian sources ($K = 2$). The k -th source of the m -th dataset, $s_k^{(m)}[n]$, is generated as

$$s_k^{(m)}[n] = (\sigma_k^{(m)} + \alpha \cdot \cos(\phi_{k,m}[n])) w_k[n] + v_k^{(m)}[n], \quad \forall k, m \in \{1, 2\}, \quad (57)$$

where

$$\phi_{k,m}[n] = \frac{2\pi}{N_{k,m}} \cdot n + \phi_0^{(k,m)}, \quad (58)$$

α is a real parameter, $\{\phi_0^{(k,m)}, N_{k,m}, \sigma_k^{(m)}\}_{k,m=1}^2$ are fixed (see Table I), and $\{w_k[n], v_k^{(m)}[n]\}_{k,m=1}^2$ are all mutually independent white standard Gaussian processes. Clearly, the k -th sources are correlated between sets, and when α is non-zero the sources are non-stationary. In the limit case where $\alpha = 0$, each source is a white Gaussian process, therefore separation cannot be attained by ICA alone (i.e., when ignoring the

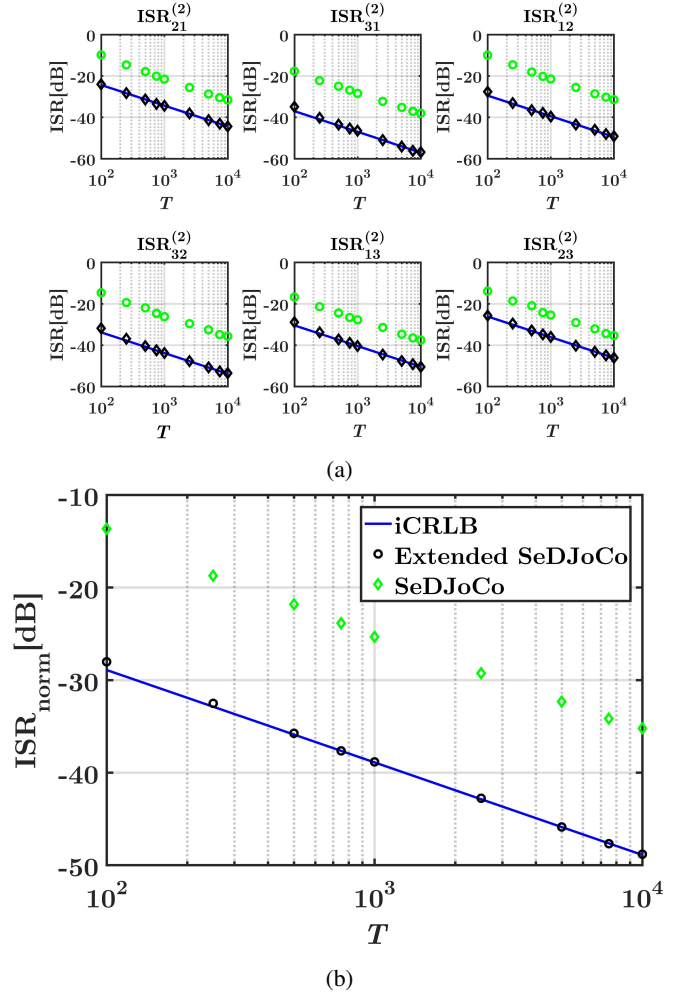


Fig. 4: (a) Empirical ISR vs. T for the second dataset (b) Empirical ISR_{norm} vs. T . The empirical results validate both the ML solution's optimality and our analytical expression of the iCRLB.

inter-datasets correlations). We compare the performance of three different solutions: the extended SeDJoCo solution (by the proposed algorithm) which yields the ML estimates w.r.t. the IVA problem altogether, the SeDJoCo solution of each dataset separately, which yields the ML estimates w.r.t. the two ICA problems separately, and Anderson *et al.*'s [26] Newton updates for Gaussian IVA (IVA-G-N), which is intended for separation of independent identically distributed (i.i.d.) Gaussian sources. In this experiment we assume that the covariance matrices of the sources are known, i.e., a semi-blind scenario, and we demonstrate how the ML solution can exploit this information in contrast to other solutions, e.g., the IVA-G-N, which cannot. The observation length was set to $T = 1000$. Fig. 3 shows all the empirical ISR elements, as well as the empirical total normalized ISR,

$$ISR_{\text{norm}} \triangleq \frac{1}{MK(K-1)} \sum_{m=1}^M \sum_{\substack{i,j=1 \\ i \neq j}}^K ISR_{ij}^{(m)}, \quad (59)$$

vs. $\alpha \in [0, 1]$, comparing also to the iCRLB derived in Section III. A good fit between the theoretical prediction and the empirical results is evident; when $\alpha = 0$ the SeDJoCo

solutions (semi-blind ICA ML approach) collapse whereas the IVA approaches give good separation in terms of the ISR. The IVA-G-N performs properly since the sources are indeed i.i.d., and the extended SeDJoCo solution (semi-blind IVA ML approach) achieves the iCRLB. As α increases, the sources become “more” non-stationary. Accordingly, IVA-G-N’s performance becomes slightly worse (due to the model mis-match) while SeDJoCo keeps improving (to the point where the cross-correlations between datasets are significantly less informative - compared to the temporal correlation of each source within each dataset). Extended SeDJoCo keeps attaining the iCRLB for all α . We stress that in this semi-blind scenario the ML solutions have an “unfair” advantage over IVA-G-N, which, unlike the ML solutions, cannot exploit the prior knowledge of the sources’ covariance matrices. Nevertheless, it is our purpose in this work to show how available prior information such as this can be exploited in an optimal manner.

Our last experiment deals with zero-lag-uncorrelated stationary sources. We consider the case where $M = K = 3$. The k -th source of the m -th dataset is generated as

$$s_k^{(m)}[n] = v_k^{(m)}[n - L \cdot (m - 1)], \quad \forall k, m \in \{1, 2, 3\}, \quad (60)$$

where

$$v_k^{(m)}[n] = \sum_{\ell=1}^M w_k^{(\ell)}[n] * h_k^{(m,\ell)}[n], \quad (61)$$

$\{w_k^{(m)}[n]\}_{k,m=1}^3$ are all mutually independent white, standard Gaussian noise processes, $\{h_k^{(m_1,m_2)}[n]\}_{k,m_1,m_2=1}^3$ are Finite Impulse Response (FIR) filters of length L for which

$$\sum_{n=0}^{L-1} |h_k^{(m_1,m_2)}[n]|^2 = \begin{cases} 1, & m_1 = m_2 \\ \eta, & m_1 \neq m_2 \end{cases}, \forall k, m_1, m_2 \in \{1, 2, 3\}, \quad (62)$$

so that η is a parameter which controls the “relative energy” contained in the cross-spectra between corresponding sources from different datasets, and $*$ denotes the convolution operator. As can be seen from (61), $h_k^{(m_1,m_2)}[n]$ is the FIR filter applied to the m_2 -th white driving-noise in order to generate a component of the k -th source in the m_1 -th dataset. Clearly, the cross-spectrum between any pair $\{s_k^{(m_1)}[n], s_k^{(m_2)}[n] : m_1 \neq m_2\}$ is non-zero when $\eta > 0$. However, note that although all such pairs are correlated, their zero-lag correlations are obviously zero (due to the $L \cdot (m - 1)$ delays).

The FIR filters were drawn from a standard Gaussian distribution with $L = 5$ and $\eta = 1$. We compare the performance of the extended SeDJoCo solution and the SeDJoCo solutions only, since the IVA-G-N algorithm requires instantaneous (zero-lag) correlation between sets and therefore performs very poorly in this scenario² (because, in addition to being zero-lag uncorrelated between sets, all sources have the same variance within sets, so they cannot even be ICA-separated using IVA-G-N, due to its inherent temporal i.i.d. model assumption). As can be seen from Fig. 4a, which shows the ISR elements of the second dataset vs. the observation length T , the SeDJoCo

solution yields quite good separation results using only the spectral diversity. However, the extended SeDJoCo solution “enjoys” not only the spectral diversity within each dataset, but also the cross-spectral diversity between the corresponding sources from different datasets. In this example, the average gain in ISR is about 15[dB] compared with the SeDJoCo solution. Similar results were obtained for the first and third datasets as well. This is reflected in Fig. 4b which shows the total normalized ISR.

VI. CONCLUSION

We presented the “extended SeDJoCo” problem, which is instrumental in finding the ML estimate of the separation matrices in the context of semi-blind IVA in a Gaussian model. This problem is also closely related to CBF in a multicast setting, and possibly to other applications. Thus, after deriving different formulations of this general problem, we outlined some of its generic properties, such as a condition for the existence of a solution and multiplicity of the solutions. We also derived two iterative solution algorithms, offering a trade-off between the required number of iterations and the computational complexity per iteration.

In the particular context of semi-blind IVA, we also derived the iCRLB on the elements of the ISR matrices for the case of Gaussian sources with arbitrary (but known) temporal auto-covariance matrices and cross-covariance matrices (between sources in different sets). We then demonstrated how this broader paradigm enables (via a solution of the extended SeDJoCo equations) the asymptotically optimal ML separation (attaining the iCRLB) of general stationary or non-stationary sources. This ability provides a significant advantage over existing IVA methods, which so far only considered the model of temporally-i.i.d. source-vector components, and moreover, could not exploit prior knowledge in a semi-blind scenario.

VII. ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support by the German-Israeli Foundation (GIF), grant number I-1282-406.10/2014. The first author also wishes to thank the Yitzhak and Chaya Weinstein Research Institute for Signal Processing for a fellowship. In addition, the authors wish to thank Tülay Adalı for providing a helpful MATLAB code of the IVA-G-N algorithm.

APPENDIX A

DIFFERENTIATION OF THE LIKELIHOOD FUNCTION

Using the following properties:

$$\begin{aligned} \text{(a)} \quad & \frac{\partial \log |\det \mathbf{X}|}{\partial \mathbf{X}} = (\mathbf{X}^{-1})^T, \text{ (b)} \quad \frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a} \mathbf{b}^T, \\ \text{(c)} \quad & \frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b} \mathbf{a}^T, \text{ (d)} \quad \mathbf{Q}_k^{(m_1, m_2)} = \mathbf{Q}_k^{(m_2, m_1)T}, \\ \text{(e)} \quad & \frac{\partial \mathbf{b}^T \mathbf{X}^T \mathbf{D} \mathbf{X} \mathbf{c}}{\partial \mathbf{X}} = \mathbf{D}^T \mathbf{X} \mathbf{b} \mathbf{c}^T + \mathbf{D} \mathbf{X} \mathbf{c} \mathbf{b}^T, \text{ (f)} \quad \mathbf{E}_{ij} \triangleq \mathbf{e}_i \mathbf{e}_j^T, \end{aligned}$$

²This was validated in simulations.

we have that

$$\begin{aligned}
\frac{\partial \mathcal{L}(\bar{\mathbf{B}})}{\partial \mathbf{B}^{(m)}} &= \frac{\partial}{\partial \mathbf{B}^{(m)}} \left(\sum_{\ell=1}^M \log |\det \mathbf{B}^{(\ell)}| \right. \\
&\quad \left. - \frac{1}{2} \sum_{k=1}^K \sum_{\substack{m_1=1 \\ m_2=1}}^M \mathbf{e}_k^T \mathbf{B}^{(m_1)} \mathbf{Q}_k^{(m_1, m_2)} \mathbf{B}^{(m_2)T} \mathbf{e}_k + \gamma \right) \\
&= \sum_{\ell=1}^M \frac{\partial \log |\det \mathbf{B}^{(\ell)}|}{\partial \mathbf{B}^{(m)}} \\
&\quad - \frac{1}{2} \sum_{k=1}^K \sum_{\substack{m_1=1 \\ m_2=1}}^M \frac{\partial \mathbf{e}_k^T \mathbf{B}^{(m_1)} \mathbf{Q}_k^{(m_1, m_2)} \mathbf{B}^{(m_2)T} \mathbf{e}_k}{\partial \mathbf{B}^{(m)}} \\
&\stackrel{(a)}{=} \mathbf{A}^{(m)T} - \frac{1}{2} \sum_{k=1}^K \sum_{\substack{m_1=1 \\ m_2=1 \\ m_2 \neq m_1}}^M \sum_{m_2=1}^M \frac{\partial \mathbf{e}_k^T \mathbf{B}^{(m_1)} \mathbf{Q}_k^{(m_1, m_2)} \mathbf{B}^{(m_2)T} \mathbf{e}_k}{\partial \mathbf{B}^{(m)}} \\
&\quad - \frac{1}{2} \sum_{k=1}^K \sum_{m_1=1}^M \frac{\partial}{\partial \mathbf{B}^{(m)}} \left(\mathbf{e}_k^T \mathbf{B}^{(m_1)} \mathbf{Q}_k^{(m_1, m_1)} \mathbf{B}^{(m_1)T} \mathbf{e}_k \right) \\
&\stackrel{(b),(c),(d),(e)}{=} \mathbf{A}^{(m)T} - \frac{1}{2} \sum_{k=1}^K 2 \sum_{m_1=1}^M \mathbf{e}_k \mathbf{e}_k^T \mathbf{B}^{(m_1)} \mathbf{Q}_k^{(m_1, m_1)} \\
&\stackrel{(f)}{=} \mathbf{A}^{(m)T} - \sum_{k=1}^K \sum_{m_1=1}^M \mathbf{E}_{kk} \mathbf{B}^{(m_1)} \mathbf{Q}_k^{(m_1, m_1)}.
\end{aligned} \tag{63}$$

REFERENCES

- [1] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. on Signal Processing*, vol. 50, no. 7, pp. 1545–1553, 2002.
- [2] A. Yeredor, "On using exact joint diagonalization for noniterative approximate joint diagonalization," *IEEE Signal Processing Letters*, vol. 12, no. 9, pp. 645–648, 2005.
- [3] J.-F. Cardoso, "Eigen-structure of the fourth-order cumulant tensor with application to the blind source separation problem," in *Proc. of ICASSP*, 1990, pp. 2655–2658.
- [4] A. Cichocki, R. Zdunek, A. H. Phan, and S.-i. Amari, *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.
- [5] X.-L. Li, T. Adali, and M. Anderson, "Joint blind source separation by generalized joint diagonalization of cumulant matrices," *Signal Processing*, vol. 91, no. 10, pp. 2314–2322, 2011.
- [6] M. Anderson, G.-S. Fu, R. Phlypo, and T. Adali, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Trans. on Signal Processing*, vol. 62, no. 17, pp. 4399–4410, 2014.
- [7] J. Via, M. Anderson, X.-L. Li, and T. Adali, "A maximum likelihood approach for independent vector analysis of Gaussian data sets," in *Proc. of MLSP*, 2011, pp. 1–6.
- [8] J. Sun, S. Papadimitriou, and S. Y. Philip, "Window-based tensor analysis on high-dimensional and multi-aspect streams," in *Proc. of ICDM*, vol. 6, 2006, pp. 1076–1080.
- [9] E. Acar, S. A. Camtepe, and B. Yener, "Collective sampling and analysis of high order tensors for chatroom communications," in *Intelligence and security informatics*. Springer, 2006, pp. 213–224.
- [10] T. G. Kolda and J. Sun, "Scalable tensor decompositions for multi-aspect data mining," in *Proc. of ICDM*, 2008, pp. 363–372.
- [11] M. Mørup, "Applications of tensor (multiway array) factorizations and decompositions in data mining," *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 24–40, 2011.
- [12] V. Stankovic and M. Haardt, "Generalized design of multi-user MIMO precoding matrices," *IEEE Trans. on Wireless Communications*, vol. 7, no. 3, pp. 953–961, 2008.
- [13] B. Song, F. Roemer, and M. Haardt, "Using a new structured joint congruence (STJOCO) transformation of Hermitian matrices for precoding in multi-user MIMO systems," in *Proc. of ICASSP*, 2010, pp. 3414–3417.
- [14] A. L. De Almeida, G. Favier, C. C. Cavalcante, and J. Mota, "Tensor-based space-time multiplexing codes for MIMO-OFDM systems with blind detection," in *Proc. of IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, 2006, pp. 1–5.
- [15] G. Favier, M. N. Da Costa, A. L. De Almeida, and J. M. T. Romano, "Tensor space-time (TST) coding for MIMO wireless communication systems," *Signal Processing*, vol. 92, no. 4, pp. 1079–1092, 2012.
- [16] J.-F. Cardoso, "Source separation using higher order moments," in *Proc. of ICASSP*, 1989, pp. 2109–2112.
- [17] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.
- [18] A.-J. Van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, 1996.
- [19] L. De Lathauwer, "A link between the canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 28, no. 3, pp. 642–666, 2006.
- [20] A. Yeredor, "On hybrid exact-approximate joint diagonalization," in *Proc. of CAMSAP*, 2009, pp. 312–315.
- [21] A. Yeredor, "Blind separation of Gaussian sources with general covariance structures: Bounds and optimal estimation," *IEEE Trans. on Signal Processing*, vol. 58, no. 10, pp. 5057–5068, 2010.
- [22] D.-T. Pham, "Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion," *Signal Processing*, vol. 81, no. 4, pp. 855–870, 2001.
- [23] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*. Academic press, 2010.
- [24] A. Yeredor, B. Song, F. Roemer, and M. Haardt, "A "sequentially drilled" joint congruence (SeDJoCo) transformation with applications in blind source separation and multiuser MIMO systems," *IEEE Trans. on Signal Processing*, vol. 60, no. 6, pp. 2744–2757, 2012.
- [25] Y.-O. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Trans. on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.
- [26] M. Anderson, T. Adali, and X.-L. Li, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Trans. on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, 2012.
- [27] J. Chatel-Goldman, M. Congedo, and R. Phlypo, "Joint BSS as a natural analysis framework for EEG-hyperscanning," in *Proc. of ICASSP*, 2013, pp. 1212–1216.
- [28] J.-H. Lee, T.-W. Lee, F. A. Jolesz, and S.-S. Yoo, "Independent vector analysis (IVA): multivariate approach for fMRI group study," *Neuroimage*, vol. 40, no. 1, pp. 86–109, 2008.
- [29] V. Calhoun, T. Adali, G. Pearlson, and J. Pekar, "A method for making group inferences from functional MRI data using independent component analysis," *Human brain mapping*, vol. 14, no. 3, pp. 140–151, 2001.
- [30] P. Banelli and S. Cacopardi, "Theoretical analysis and performance of OFDM signals in nonlinear AWGN channels," *IEEE Trans. on Communication*, vol. 48, no. 3, pp. 430–441, 2000.
- [31] Y. Cheng, A. Yeredor, A. Weiss, and M. Haardt, "Extension of the "sequentially drilled" joint congruence transformation (SeDJoCo) problem," in *Proc. of CAMSAP*, 2015, pp. 185–188.
- [32] C. G. Gibson, *Elementary Geometry of Algebraic Curves: An Undergraduate Introduction*. Cambridge University Press, 2001.
- [33] A. Yeredor, Y. Cheng, and M. Haardt, "On multiple solutions of the "sequentially drilled" joint congruence transformation (SeDJoCo) problem for semi-blind source separation," in *Proc. of ICASSP*, 2016.
- [34] A. Weiss, A. Yeredor, S. A. Cheema, and M. Haardt, "Maximum likelihood "identification-correction" scheme of sub-optimal "SeDJoCo" solutions for Gaussian blind source separation," in *Proc. of ICASSP*, 2017.
- [35] E. Doron, A. Yeredor, and P. Tichavsky, "Cramér-Rao-induced bound for blind separation of stationary parametric Gaussian sources," *IEEE Signal Processing Letters*, vol. 14, no. 6, pp. 417–420, 2007.
- [36] Y. Cheng, A. Yeredor, and M. Haardt, "Extension of SeDJoCo and its use in a combination of multicast and coordinated multi-point systems," in *Proc. of ICASSP*, 2016, pp. 3276–3280.
- [37] S. Dégerine and A. Zaïdi, "Separation of an instantaneous mixture of Gaussian autoregressive sources by the exact maximum likelihood approach," *IEEE Trans. on Signal Processing*, vol. 52, no. 6, pp. 1499–1512, 2004.
- [38] K. B. Petersen, M. S. Pedersen *et al.*, "The matrix cookbook," *Technical University of Denmark*, vol. 7, p. 15, 2008.