PERFORMANCE BOUND FOR BLIND EXTRACTION OF NON-GAUSSIAN COMPLEX-VALUED VECTOR COMPONENT FROM GAUSSIAN BACKGROUND

Václav Kautský[1], *Zbyněk Koldovský*[2] *and Petr Tichavský*[3]

[1]Faculty of Nuclear Sciences and Physical Engineering,
Czech Technical University in Prague, E-mail: kautsvac@fjfi.cvut.cz
[2]Faculty of Mechatronic and Interdisciplinary Studies,
Technical University of Liberec, Studentská 2, 461 17 Liberec, Czech Republic
[3]The Czech Academy of Sciences, Institute of Information Theory and Automation,
P. O. Box 18, 182 08 Prague 8, Czech Republic

ABSTRACT

Independent Vector Extraction aims at the joint blind source extraction of K dependent signals of interest (SOI) from K mixtures (one signal from one mixture). Similarly to Independent Component/Vector Analysis (ICA/IVA), the SOIs are assumed to be independent of the other signals in the mixture. Compared to IVA, the (de-)mixing IVE model is reduced in the number of parameters for the extraction problem. The SOIs are assumed to be non-Gaussian or noncircular Gaussian, while the other signals are modeled as circular Gaussian. In this paper, a Cramér-Rao-Induced Bound (CRIB) for the achievable Interference-to-Signal Ratio (ISR) is derived for IVE. The bound is compared with similar bounds for ICA, IVA, and Independent Component Extraction (ICE). Numerical simulations show a good correspondence between the empirical results and the theory.

Index Terms— Blind Source Extraction, Independent Component Analysis, Independent Vector Analysis, Cramér-Rao Bound

1. INTRODUCTION

Independent Vector Analysis (IVA) aims at simultaneous separation of $K > 1$ linear mixtures of independent signals [1]. It is a generalization of Independent Component Analysis (ICA) where only one mixture of scalar sources (components) is separated [2]. IVA is formulated so that vector components are separated where a vector component is formed from K scalar, possibly dependent, sources, one source from each mixture. The key assumption for the separability is the mutual independence of the (vector) components [3, 4].

Specifically, let the kth mixture, $k = 1, \dots, K$, be described by $\mathbf{x}^k = \mathbf{A}^k \mathbf{s}^k$ where $\mathbf{s}^k \in \mathbb{C}^{d\times 1}$ is the vector of independent complex-valued source signals (scalar random variables), $\mathbf{A}^k \in \mathbb{C}^{d\times d}$ is the unknown mixing matrix, and $\mathbf{x}^k \in \mathbb{R}^{d\times 1}$ is the vector of the observed signals. The joint IVA mixing model can be written as $\mathbf{x} = \mathbf{A}\widetilde{\mathbf{s}}$, where $\widetilde{\mathbf{s}} = [\mathbf{s}^1; \dots; \mathbf{s}^K] \in \mathbb{C}^{dK\times 1}$, $\mathbf{x} = [\mathbf{x}^1; \dots; \mathbf{x}^K]$, and $\mathbf{A}$ is block-diagonal with the kth diagonal block equal to $\mathbf{A}^k$.

The jth vector component, $j = 1, \dots, d$, is $\mathbf{s}_j = [s_j^1; \dots; s_j^K]^T \in \mathbb{R}^{K\times 1}$. The elements within each $\mathbf{s}_j$ can be dependent, however, elements from different vector components are assumed to be mutually independent. Here, we will focus on the IVA problem where the elements of the vector components are uncorrelated, so only higher-order dependencies among them can exist [5]. Separation of vector components with internal correlations can be based on using second-order statistics only; see, e.g., [6, 7, 8].

This paper is focused on bounds for the achievable separation accuracy. In particular, we consider a Blind Source Extraction (BSE) problem where only one vector component should be extracted. We consider BSE performed through the approach referred to as Independent Vector Extraction (IVE), which is generalization of Independent Component Extraction (ICE) [9, 10]. In ICE/IVE, the mixing models $\mathbf{x}^k = \mathbf{A}^k \mathbf{s}^k$ are re-parameterized for BSE. The mixing matrices are structured where each one is parameterized only by two (mixing and separating) vectors. It was shown that algorithms derived based on ICE/IVE are closely related to previous well-known BSE methods such as One-unit FastICA [11] or to the Natural Gradient algorithm [12, 1]. Moreover, Cramér-Rao Lower Bound (CRLB) for ICE (the real-valued case studied in [13]) shows that the achievable accuracy through ICE is in agreement with the asymptotic accuracy of One-unit FastICA [14, 15]. Consequently, the reduction of the mixing model does not cause any performance limitations and is therefore useful for further development of BSE methods.

In this paper, we analyze the accuracy limitations for the complex-valued IVE problem. A novel CRLB-based bound for the achievable Interference-to-Signal Ratio (ISR) is derived. The vector source of interest (SOI) is assumed to have non-Gaussian or Gaussian non-circular elements while the other interfering sources (referred to as background) are assumed to be circular Gaussian. The bound is verified by simulations and compared with that derived for IVA [4], where all sources are assumed to be non-Gaussian. A related bound for ICE is also compared, showing clear advantage of the joint extraction compared to the non-joint one.

2. PROBLEM FORMULATION

2.1. Reduced Mixing Model

In general, the kth mixture can be described as $\mathbf{x}^k = \mathbf{a}^k s^k + \mathbf{y}^k$, where s^k, $\mathbf{a}^k$, and $\mathbf{y}^k$ is the SOI, the mixing vector corresponding to the SOI, and the background sources, respectively. This model is suitable when only the SOI should be extracted. However, to derive the CRLB for the extraction problem, the mixing model must be invertible so that the probability density of $\mathbf{x}^k$ can be expressed.

[0]This work was partly supported by The Czech Science Foundation through Project No. 17-00902S and by the United States Department of the Navy, Office of Naval Research Global, through Project No. N62909-18-1-2040 .

Therefore, we proceed by defining the inverse of the mixing system as follows. Let $\mathbf{w}^k$ be the separating vector such that $s^k = (\mathbf{w}^k)^H\mathbf{x}^k$. Then, the de-mixing matrix can have the form

$$\mathbf{W}_{\rm ICE}^k = \begin{pmatrix} {\mathbf{w}^k}^H \\ \mathbf{B}^k \end{pmatrix} = \begin{pmatrix} \beta^{k*} & {\mathbf{h}^k}^H \\ \mathbf{g}^k & -\gamma^k\mathbf{I}_{d-1} \end{pmatrix}, \tag{1}$$

where γ^k and $\mathbf{g}^k$ denote the upper and lower parts of $\mathbf{a}^k$, respectively, i.e., $\mathbf{a}^k = [\gamma^k; \mathbf{g}^k]$, and similarly $\mathbf{w}^k = [\beta^k; \mathbf{h}^k]$. $\mathbf{W}_{\rm ICE}^k$ separates $\mathbf{x}^k$ into two components $\mathbf{v}^k = \mathbf{W}^k\mathbf{x}^k = [s^k; \mathbf{z}^k]$ where $\mathbf{z}^k = \mathbf{B}^k\mathbf{x}^k$ does not contain s^k, because $\mathbf{B}^k\mathbf{a}^k = \mathbf{0}$. Hence, the mixing model is

$$\mathbf{x}^k = \mathbf{A}_{\rm ICE}^k\mathbf{v}^k, \tag{2}$$

where $\mathbf{A}_{\rm ICE}^k = (\mathbf{W}_{\rm ICE}^k)^{-1}$. By taking into account (1),

$$\mathbf{A}_{\rm ICE}^k = [\mathbf{a}^k, \quad \mathbf{Q}^k] = \begin{pmatrix} \gamma^k & {\mathbf{h}^k}^H \\ \mathbf{g}^k & \frac{1}{\gamma^k}\left(\mathbf{g}^k{\mathbf{h}^k}^H - \mathbf{I}_{d-1}\right) \end{pmatrix}, \tag{3}$$

where β^k and γ^k satisfy $\beta^k\gamma^k = 1 - {\mathbf{h}^k}^H\mathbf{g}^k$.

Compared to the original ICA mixing model $\mathbf{x}^k = \mathbf{A}^k\mathbf{s}^k$ with the fully parametrized $\mathbf{A}^k$, $\mathbf{A}_{\rm ICE}^k$ depends purely on $\mathbf{w}^k$ and $\mathbf{a}^k$, where $\mathbf{a}^k$ corresponds to any column of $\mathbf{A}^k$ (depending on which source in $\mathbf{s}^k$ is the SOI). Without any loss of generality, let the SOI be the first source, i.e., $s^k = s_1^k$; the vector source of interest corresponds to $\mathbf{s} = \mathbf{s}_1$.

Also, (2) corresponds with the description $\mathbf{x}^k = \mathbf{a}^k s^k + \mathbf{y}^k$ provided that $\mathbf{y}^k = \mathbf{Q}^k\mathbf{z}^k$. This means that in (2) it is inherently assumed that $\mathbf{y}^k$ is generated from $d-1$ latent sources $\mathbf{z}^k$.

The joint mixture model for the extraction of $\mathbf{s}_1$ can be written as $\mathbf{x} = \mathbf{A}_{\rm IVE}\mathbf{v}$, where $\mathbf{v} = [\mathbf{v}^1; \dots; \mathbf{v}^K]$ and $\mathbf{A}_{\rm IVE}$ is block diagonal with the kth block equal to $\mathbf{A}_{\rm ICE}^k$. One goal of this paper is to show that this parameter-reduced model is suitable for the BSE problem and does not bring any restriction compared to the fully parameterized ICA model.

2.2. Independence-based probability model with circularly Gaussian background

Similarly to IVA, the fundamental assumption is that $\mathbf{s}$ and $\mathbf{z} = [\mathbf{z}^1; \dots; \mathbf{z}^K]$ are mutually independent, so their joint pdf is a product of the pdf of the target signal, $p_{\mathbf{s}}(\mathbf{s})$, and the pdf of the background $p_{\mathbf{z}}(\mathbf{z})$. The fact that the pdfs of $\mathbf{s}$ and $\mathbf{z}$ are not further factorized into products of marginals enables the internal dependence, which is the main IVE concept adopted from IVA. Now, by exploiting (1), the joint pdf of the observation vector $\mathbf{x}$ reads

$$p(\mathbf{x}|\mathbf{a}, \mathbf{w}) = p_{\mathbf{s}}(\{{\mathbf{w}^k}^H\mathbf{x}^k\}_{k=1}^K)p_{\mathbf{z}}(\{\mathbf{B}^k\mathbf{x}^k\}_{k=1}^K)|\det \mathbf{A}_{\rm IVE}^{-1}|^2 \tag{4}$$

This paper is focused on the particular case when $\mathbf{z}$ are circular Gaussian, i.e. $\mathbf{z} \sim \mathcal{CN}(0, \mathbf{C}_{\mathbf{z}})$, where $\mathbf{C}_{\mathbf{z}} = \mathrm{E}\left[\mathbf{z}\mathbf{z}^H\right]$. Since $\mathbf{z}^1, \dots, \mathbf{z}^K$ are assumed to be uncorrelated, $\mathbf{C}_{\mathbf{z}}$ is block diagonal with the kth diagonal block equal to $\mathbf{C}_{\mathbf{z}}^k = \mathrm{E}[\mathbf{z}^k{\mathbf{z}^k}^H]$; let's denote $\mathbf{R}^k = {\mathbf{C}_{\mathbf{z}}^k}^{-1}$.

The scaling ambiguity can be fixed by putting $\gamma^k = 1$, and then $\det(\mathbf{W}_{\rm ICE}^k) = 1$. The remaining free parameters of the model are represented by $\mathbf{g}^k$ and $\mathbf{h}^k$, $k = 1, \dots, K$, and the log-pdf of the observed data takes the form

$$\mathcal{L}(\mathbf{X}|\mathbf{g}, \mathbf{h}) = \log p_{\mathbf{s}}(\{{\mathbf{w}^k}^H\mathbf{x}^k\}_{k=1}^K) + \sum_{k=1}^K \log(\det \mathbf{R}^k) - \\ - \sum_{k=1}^K {\mathbf{x}^k}^H{\mathbf{B}^k}^H\mathbf{R}^k\mathbf{B}^k\mathbf{x}^k - K(d-1)\log(2\pi). \tag{5}$$

3. CRLB FOR IVE

In the following, we will use the CRLB definition for complex-valued parameters from [16, 17]. Let $\boldsymbol{\theta}^k = [\mathbf{g}^k; \mathbf{h}^k]$ denote the parameter vector for the kth mixture, $\boldsymbol{\theta} = [\boldsymbol{\theta}^1; \dots; \boldsymbol{\theta}^K]$, and $\tilde{\boldsymbol{\theta}} = [\boldsymbol{\theta}; \boldsymbol{\theta}^*]$. For any unbiased estimator of $\tilde{\boldsymbol{\theta}}$, it holds that [17]

$$\mathtt{cov}\left(\tilde{\boldsymbol{\theta}}\right) \succeq \mathcal{J}^{-1}(\tilde{\boldsymbol{\theta}}) = \mathtt{CRLB}(\tilde{\boldsymbol{\theta}}), \tag{6}$$

where $\mathbf{C} \succeq \mathbf{D}$ means that $\mathbf{C} - \mathbf{D}$ is positive semi-definite, and $\mathcal{J}(\tilde{\boldsymbol{\theta}})$ is the Fisher information matrix (FIM) defined (in a block structure) as

$$\mathcal{J}(\tilde{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{F} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{F}^* \end{pmatrix} = \mathrm{E}\left[\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}}\right)^H\right], \tag{7}$$

where

$$\mathbf{F} = \mathrm{E}\left[\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}^*}\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}^*}\right)^H\right], \quad \mathbf{P} = \mathrm{E}\left[\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}^*}\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}^*}\right)^T\right], \tag{8}$$

The derivatives with respect to $\boldsymbol{\theta}^*$ are defined according to the Wirtinger calculus [16, 18].

3.1. Lower Bound for Interference-to-Signal Ratio

Let $\widehat{\mathbf{w}}^k$ denote an estimated separating vector for the kth mixture. The Interference-to-Signal Ratio (ISR) achieved by the estimator is defined as the variance of the interference divided by the variance of the SOI

$$\mathtt{ISR}(\widehat{\mathbf{w}}^k) = \frac{(\widehat{\mathbf{w}}^k)^H\mathbf{C}_{\mathbf{y}}^k\widehat{\mathbf{w}}^k}{\sigma_{s^k}^2|(\widehat{\mathbf{w}}^k)^H\mathbf{a}^k|^2} = \frac{(\widehat{\mathbf{q}}_2^k)^H\mathbf{C}_{\mathbf{z}}^k\widehat{\mathbf{q}}_2^k}{|\widehat{q}_1^k|^2\sigma_{s^k}^2} \approx \frac{(\widehat{\mathbf{q}}_2^k)^H\mathbf{C}_{\mathbf{z}}^k\widehat{\mathbf{q}}_2^k}{\sigma_{s^k}^2}, \tag{9}$$

where $\mathbf{C}_{\mathbf{y}}^k = \mathrm{E}[\mathbf{y}^k\mathbf{y}^k]^H$, $(\widehat{\mathbf{q}}^k)^T = [\widehat{q}_1^k, (\widehat{\mathbf{q}}_2^k)^T] = (\widehat{\mathbf{w}}^k)^H\mathbf{A}_{\rm ICE}^k = \left[(\widehat{\mathbf{w}}^k)^H\mathbf{a}^k, (\widehat{\mathbf{w}}^k)^H\mathbf{Q}^k\right]$. The last approximation in (9) assumes "small" errors, i.e. $(\widehat{q}_1^k)^2 \approx 1$ and $\widehat{\mathbf{q}}^k \approx \mathbf{e}_1$ (the unit vector). Then, the mean ISR value reads

$$\mathrm{E}[\mathtt{ISR}(\widehat{\mathbf{w}}^k)] \approx \frac{\mathrm{E}\left[(\widehat{\mathbf{q}}_2^k)^H\mathbf{C}_{\mathbf{z}}^k\widehat{\mathbf{q}}_2^k\right]}{\sigma_{s^k}^2} = \frac{\mathtt{tr}\left(\mathbf{C}_{\mathbf{z}}^k\mathrm{E}[\widehat{\mathbf{q}}_2^k(\widehat{\mathbf{q}}_2^k)^H]\right)}{\sigma_{s^k}^2}. \tag{10}$$

Owing to the equivariance property of the BSE problem [2], we can consider the special case when $\mathbf{h} = \mathbf{0}$. Then, $\widehat{\mathbf{q}}_2^k = \widehat{\mathbf{h}}^k$, and

$$\mathrm{E}[\mathtt{ISR}(\widehat{\mathbf{w}}^k)] \approx \frac{\mathtt{tr}\left(\mathbf{C}_{\mathbf{z}}^k\mathtt{cov}(\widehat{\mathbf{q}}_2^k)\right)}{\sigma_{s^k}^2} = \frac{\mathtt{tr}\left(\mathbf{C}_{\mathbf{z}}^k\mathtt{cov}(\widehat{\mathbf{h}}^k)\right)}{\sigma_{s^k}^2}, \tag{11}$$

then

$$\mathrm{E}[\mathtt{ISR}(\widehat{\mathbf{w}}^k)] \geq \sigma_{s^k}^{-2}\mathtt{tr}\left(\mathbf{C}_{\mathbf{z}}^k\mathtt{CRLB}(\mathbf{h}^k)\right), \tag{12}$$

where $\mathtt{CRLB}(\mathbf{h}^k)$ denotes the kth diagonal block of $\mathcal{J}^{-1}(\tilde{\boldsymbol{\theta}})$ corresponding to the parameter $\mathbf{h}^k$. As shown in Appendix A, the FIM (7) is equal to

$$\mathcal{J}(\tilde{\boldsymbol{\theta}}) = \begin{pmatrix} \mathbf{F} & \mathbf{O} \\ \mathbf{O} & \mathbf{F}^* \end{pmatrix}, \text{ where } \mathbf{F} = \begin{pmatrix} \mathbf{F}_{\mathbf{g},\mathbf{g}} & -\mathbf{I}_{d-1} \\ -\mathbf{I}_{d-1} & \mathbf{F}_{\mathbf{h},\mathbf{h}} \end{pmatrix}. \tag{13}$$

Using the block-structure of (13) for computing $\mathcal{J}^{-1}(\tilde{\boldsymbol{\theta}})$, the CRLB reads

$$\mathtt{CRLB}(\mathbf{h}^k) = \mathbf{R}^k \frac{\sigma_{s^k}^2}{\kappa_{\mathrm{IVE}}^k \sigma_{s^k}^2 - 1}. \tag{14}$$

Using (12) and (14), and by considering N observations, the CRLB-induced bound for ISR for the kth mixture is

$$\mathrm{E}[\mathtt{ISR}(\widehat{\mathbf{w}}^k)] \geq \frac{1}{N} \frac{d-1}{\kappa_{\mathrm{IVE}}^k \sigma_{s^k}^2 - 1}. \tag{15}$$

It can be shown, that $\kappa_{\mathrm{IVE}}^k \sigma_{s^k}^2 = \tilde{\kappa}_{\mathrm{IVE}}^k$, where $\tilde{\kappa}_{\mathrm{IVE}}^k$ corresponds to the same pdf as κ_{IVE}^k but scaled to the unit variance. Thus, the resulting bound does not depend on the variance of original signals.

3.2. Resulting bound vs. known bounds for ICA/IVA

Now, we compare the result (15) with previous known bounds for IVA, ICA and ICE.

The CRIB for IVA derived in [4] (only the real-valued case) says that the ISR for the mth component satisfies

$$\mathrm{E}[\mathtt{ISR}(\widehat{\mathbf{w}}_m^k)] \geq \sum_{\substack{n=1 \\ n \neq m}}^{d} \frac{1}{N} \left(\left(\boldsymbol{\nu}_{m,n} - \boldsymbol{\nu}_{n,m}^{-1} \right)^{-1} \right)_{kk}, \tag{16}$$

where $\boldsymbol{\nu}_{m,n} = \mathrm{E}[\boldsymbol{\psi}_m(\mathbf{s}_m)\boldsymbol{\psi}_m^T(\mathbf{s}_m)] \odot \mathrm{E}\left[\mathbf{s}_n\mathbf{s}_n^T\right]$, $n, m = 1, \ldots d$, and $\boldsymbol{\psi}_m(\mathbf{s}_m) = -\partial \log p_m(\mathbf{s}_m)/\partial \mathbf{s}_m$; $\odot$ denotes the Hadamard (element-wise) product. To compare (16) with (15), consider $m = 1$. Next, the elements of vector components are assumed to be uncorrelated, so $\mathrm{E}\left[\mathbf{s}_n\mathbf{s}_n^T\right]$ and consequently $\boldsymbol{\nu}_{m,n}$ are diagonal. Moreover, when the nth vector component is Gaussian, $\boldsymbol{\nu}_{n,1} = \left(\mathrm{E}\left[\mathbf{s}_n\mathbf{s}_n^T\right]\right)^{-1} \odot \mathrm{E}\left[\mathbf{s}_1\mathbf{s}_1^T\right]$. Using these simplifications, it can be verified that (16) coincides with (15) when all but the first vector components are Gaussian, which corresponds with the assumption introduced in Section 2.2.

However, the IVA mixing model is not identifiable under the assumption of Gaussian background [4], consequently, the corresponding CRLB formally does not exist. By contrast, the CRLB for IVE exists, and (15) corresponds with the asymptotic value of (16) for $m = 1$. In conclusion, the reduced mixing model in Section 2.1 does not limit the achievable accuracy by IVE compared to IVA. Only when the background signals are non-Gaussian, (16) can be lower than (15). Similar conclusions hold when comparing ICE with ICA [13].

ICE differs from IVE in that each mixture is treated separately. Then, the CRIB for the kth mixture has the form [13]

$$\mathrm{E}[\mathtt{ISR}(\widehat{\mathbf{w}}^k)] \geq \frac{1}{N} \frac{d-1}{\kappa_{\mathrm{ICE}}^k \sigma_{s^k}^2 - 1}, \tag{17}$$

where $\kappa_{\mathrm{ICE}}^k = \mathrm{E}[(\partial p_k(s^k)/\partial s^k)^2]$, and $p_k(s^k)$ stands for the marginal pdf of s^k.

By comparing, (17) and (15) differ only in that κ^k depends on the score function of s^k derived from its marginal and joint pdf, respectively.

Proposition 1. *Let $p(s^1, \ldots, s^K)$ denote the joint pdf of $s^1, \ldots, s^K$, and $p_k(s^k)$ be the marginal pdf of s^k, $k = 1, \ldots, K$. Then, $\kappa_{\mathrm{IVE}}^k \geq \kappa_{\mathrm{ICE}}^k$, and the equality when s^k is independent of the other random variables, or, equivalently, when $p(s^1, \ldots, s^K) = p_k(s^k)p(s_1, \ldots, s^{k-1}, s^{k+1}, \ldots, s^K)$.*

The Proposition 1 (see Appendix B for the proof) shows that the CRIB for ICE is always greater than for IVE. Once s^k is independent of the other elements of $\mathbf{s}$, the CRIBs are the same. It holds that $\kappa_{\mathrm{ICA}}^k = \kappa_{\mathrm{IVA}}^k$ if and only if s^k is independent of the other elements of $\mathbf{s}$. IVE thus takes advantage compared to ICE only when there are dependencies inside the vector component $\mathbf{s}$. The equality in comparison of ICA vs. ICE and IVA vs. IVE holds when all but one signals in a mixture are Gaussian. Finally, Table 1 provides a summary of inequalities between the CRIBs.

	ICA	ICE	IVA	IVE
ICA	=	$\leq$	$\geq$	n/a
ICE	$\geq$	=	n/a	$\geq$
IVA	$\leq$	n/a	=	$\leq$
IVE	n/a	$\leq$	$\geq$	=

Table 1. Comparison of CRIBs for $\mathrm{E}[\mathtt{ISR}(\widehat{\mathbf{w}}^k)]$.

4. SIMULATIONS

In simulations, we compare the bounds for ICE and IVE with empirical mean ISR achieved by the OGICE (Orthogonally Constrained ICE) algorithm from [9], and by OGIVE performing IVE [10]. Both algorithms are based on maximum likelihood principle, so they might achieve the CRIB asymptotically when initialized in the region of convergence to the SOI, and when the true score function is used as the internal nonlinear function. For simplicity we assume real-valued signals and mixing matrix.

In one trial, $K = 3$ mixtures of $d = 5$ independent signals are generated. The background signals in mixtures are Gaussian with zero mean and unit variance. The SOIs (one SOI per mixture) are mutually dependent, drawn according to the joint pdf given by

$$p(s^1, \ldots, s^K) \propto \exp\left(- \left(\lambda \sum_{i=1}^{K} |s^i|^2 \right)^{\alpha} \right) \tag{18}$$

where $\lambda > 0$, and $\alpha \neq 1$ (for $\alpha = 1$, the pdf is Gaussian). To scale the marginal pdf of the SOI for $K = 3$ mixtures to the unit variance we set $\lambda = \frac{\Gamma\left(\frac{5}{2\alpha}\right)}{3\Gamma\left(\frac{3}{2\alpha}\right)}$. The SOIs are generated by using the Markov Chain Monte Carlo sampler [19]. All signals are mixed by a random mixing matrix.

For $K = 3$, the marginal pdf of the kth SOI (assuming zero mean and unit variance) is given by

$$p_k(s^k) = D_2 \int_{(\lambda |s^k|^2)^\alpha} t^{\frac{1}{\alpha}-1} \mathrm{e}^{-t} \mathrm{d}t = D_2 \Gamma\left((\lambda |s^k|^2)^\alpha, \frac{1}{\alpha} \right), \tag{19}$$

where $D_2 = \frac{\sqrt{\lambda}}{2\Gamma\left(\frac{3}{2\alpha}\right)}$, $\Gamma(y, a)$ denotes the incomplete Gamma function. In OGIVE, the nonlinearity is the true score function derived from the joint pdf (18), i.e.,

$$\psi_{\mathrm{IVE}}^k = -\frac{\partial \log p(s^1, \ldots, s^K)}{\partial s^k} = 2\alpha\lambda^\alpha s^k \left(\sum_{i=1}^{3} |s^i|^2 \right)^{\alpha - 1}, \tag{20}$$

while, in OGICE, the score function corresponding to the marginal pdf (19) is used, that is,

$$\psi^k_{\text{ICE}} = -\frac{\partial \log p_k(s^k)}{\partial s^k} = \frac{2\alpha\lambda s^k \exp\left(-(\lambda|s^k|^2)^\alpha\right)}{\Gamma\left((\lambda|s^k|^2)^\alpha, \frac{1}{\alpha}\right)}. \tag{21}$$

It holds that

$$\kappa^k_{\text{IVE}} = \mathrm{E}\left[{\psi^k_{\text{IVE}}}^2\right] = \frac{4}{3}\lambda\alpha^2 \frac{\Gamma(2+\frac{1}{2\alpha})}{\Gamma(\frac{3}{2\alpha})}. \tag{22}$$

The values of κ^k_{ICE} were computed numerically.
Fig. 4 shows the CRIBs given by (15) and (17) together with the average ISR achieved by the algorithms in 100 trials, as functions of $\alpha \in [0.2, 2]$. The results confirm that the CRIB for IVE is smaller than the CRIB for ICE when the SOIs are dependent. Also, the empirical results by OGICE and OGIVE confirm that the latter achieves lower ISR. The average ISRs are close to their respective CRIBs, but they are always slightly greater.
The CRIBs and the average ISRs grow when α is close one, because the pdf of the SOI is close to Gaussian. For the Gaussian case, the mixing model is not identifiable. For α close to one, the estimator seems to perform even slightly better than the CRLB, but it is only because of its initialization in the true value of $\mathbf{g}$.

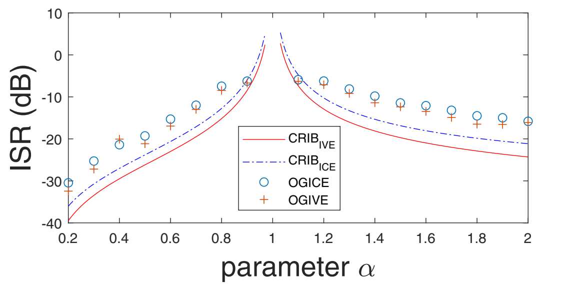

Fig. 1. CRIBs and average ISRs achieved by the compared algorithms for $d = 5$, $N = 5000$, $K = 3$.

5. CONCLUSIONS

The CRIB on ISR achieved by IVE derived in this paper has shown that the structured (de-)mixing matrix model with a reduced number of parameters is not restrictive. The accuracy achievable by IVE is, in comparison to IVA, asymptotically the same when the background is Gaussian. The dependence between the SOIs in the mixtures enable IVE to reach better accuracy than ICE, which treats each mixture separately. Numerical simulations have confirmed the validity of the CRIBs.

Appendix A: Fisher Information Matrix

Now, we complete the CRLB-induced bound for ISR by computing the FIM and its inverse. Let $\mathbf{F}$ defined in (8) be partitioned as

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{\mathbf{g},\mathbf{g}} & \mathbf{F}_{\mathbf{g},\mathbf{h}} \\ \mathbf{F}^*_{\mathbf{g},\mathbf{h}} & \mathbf{F}_{\mathbf{h},\mathbf{h}} \end{pmatrix}, \text{ where } \mathbf{g} = [\mathbf{g}^1; \ldots; \mathbf{g}^K] \text{ and } \mathbf{h} = [\mathbf{h}^1; \ldots; \mathbf{h}^K]. \tag{23}$$

The derivatives of (5) read

$$\frac{\partial \mathcal{L}(\mathbf{x}^1 \ldots \mathbf{x}^K|\mathbf{g},\mathbf{h})}{\partial \mathbf{g}^k}\Big|_{\mathbf{h}=\mathbf{0}} = -s^k \mathbf{R}^k \mathbf{z}^k \quad \text{and} \quad \frac{\partial \mathcal{L}(\mathbf{x}^1 \ldots \mathbf{x}^K|\mathbf{g},\mathbf{h})}{\partial \mathbf{h}^k}\Big|_{\mathbf{h}=\mathbf{0}} = \psi^k(s^1, \ldots, s^K)\mathbf{z}^k, \tag{24}$$

where $\psi^k(s^1, \ldots, s^K) = -\frac{\partial \log p(s^1,\ldots,s^K)}{\partial s^k}$. The elements of the kth diagonal block of (23) are as follows (the off-diagonal blocks are zero since $\mathbf{z}^v$, $\mathbf{z}^u$, $u \neq v$, are uncorrelated): $(\mathbf{F}_{\mathbf{g},\mathbf{g}})^k = \sigma^2_{s^k}\mathbf{R}^k$, $(\mathbf{F}_{\mathbf{h},\mathbf{h}})^k = \kappa^k_{\text{IVE}}\mathbf{C}^k_{\mathbf{z}}$, $(\mathbf{F}_{\mathbf{g},\mathbf{h}})^k = -\mathbf{I}_{d-1}$, where $\kappa^k_{\text{IVE}} = \mathrm{E}[|\psi^k(\mathbf{s})|^2]$. The other sub-matrices of (7) can be partitioned similarly as (23). Then, by taking into account the circularity of $\mathbf{z}^k$, which means that $\mathrm{E}[\mathbf{z}^k {\mathbf{z}^k}^T] = \mathbf{O}$, it follows that $(\mathbf{P}_{\mathbf{g},\mathbf{g}})^k = (\mathbf{P}_{\mathbf{h},\mathbf{h}})^k = (\mathbf{P}_{\mathbf{g},\mathbf{h}})^k = \mathbf{O}$.

Appendix B: Proof of $\kappa^k_{\text{IVE}} \geq \kappa^k_{\text{ICE}}$

Proof: We start by computing the following auxiliary quantity:

$$\kappa^k_{\text{MI}} = \mathrm{E}\left[\left(\frac{\partial}{\partial s^k}\left(\log \frac{p(s^1,\ldots,s^K)}{\prod_{i=1}^K p_i(s^i)}\right)\right)^2\right] = \mathrm{E}\left[\left(\frac{\partial \log p(s^1,\ldots,s^K)}{\partial s^k}\right)^2\right] + \mathrm{E}\left[\left(\frac{\partial \log\left(\prod_{i=1}^K p_i(s^i)\right)}{\partial s^k}\right)^2\right] -$$

$$2\mathrm{E}\left[\frac{\partial \log p(s^1,\ldots,s^K)}{\partial s^k}\frac{\partial \log\left(\prod_{i=1}^K p_i(s^i)\right)}{\partial s^k}\right] = \kappa^k_{\text{IVE}} + \kappa^k_{\text{ICE}} - 2\mathrm{E}\left[\frac{\partial \log p(s^1,\ldots,s^K)}{\partial s^k}\frac{\partial \log p_k(s^k)}{\partial s^k}\right].$$

By unfolding the last term, we obtain

$$\int_{\mathbf{R}^K} \frac{1}{p(s^1,\ldots,s^K)}\frac{\partial p(s^1,\ldots,s^K)}{\partial s^k}\frac{1}{p_k(s^k)}\frac{\partial p_k(s^k)}{\partial s^k} p(s^1,\ldots,s^K)\mathtt{d}s^1\ldots\mathtt{d}s^K =$$

$$\int_{\mathbf{R}} \frac{1}{p_k(s^k)}\frac{\partial p_k(s^k)}{\partial s^k}\int_{\mathbf{R}^{K-1}}\frac{\partial p(s^1,\ldots,s^K)}{\partial s^k}\mathtt{d}s^1\ldots\mathtt{d}s^K = \int_{\mathbf{R}}\frac{1}{p_k(s^k)}\left(\frac{\partial p_k(s^k)}{\partial s^k}\right)^2\mathtt{d}s^k = \kappa^k_{\text{ICE}}.$$

The proof is completed as $0 \leq \kappa^k_{\text{MI}} = \kappa^k_{\text{IVE}} - \kappa^k_{\text{ICE}}$. ■

6. REFERENCES

[1] T. Kim, I. Lee, and T. Lee, "Independent vector analysis: Definition and algorithms," in *2006 Fortieth Asilomar Conference on Signals, Systems and Computers*, Oct 2006, pp. 1393–1396.

[2] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, ser. Independent Component Analysis and Applications Series. Elsevier Science, 2010.

[3] J. Eriksson and V. Koivunen, "Identifiability, separability, and uniqueness of linear ICA models," *IEEE Signal Processing Letters*, vol. 11, no. 7, pp. 601–604, July 2004.

[4] M. Anderson, G. Fu, R. Phlypo, and T. Adalı, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4399–4410, Sept 2014.

[5] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, pp. 70–79, Jan. 2007.

[6] Y. 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.

[7] D. Lahat and C. Jutten, "Joint independent subspace analysis using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4891–4904, Sept 2016.

[8] A. Weiss, S. A. Cheema, M. Haardt, and A. Yeredor, "Performance analysis of the gaussian quasi-maximum likelihood approach for independent vector analysis," *IEEE Transactions on Signal Processing*, vol. 66, no. 19, pp. 5000–5013, Oct 2018.

[9] Z. Koldovský, P. Tichavský, and V. Kautský, "Orthogonally constrained independent component extraction: Blind MPDR beamforming," in *Proceedings of European Signal Processing Conference*, Sep. 2017, pp. 1195–1199.

[10] Z. Koldovský and P. Tichavský, "Gradient algorithms for complex non-gaussian independent component/vector extraction, question of convergence," *IEEE Transactions on Signal Processing*, vol. 67, no. 4, pp. 1050–1064, Feb 2019.

[11] A. Hyvärinen, "Fast and robust fixed-point algorithm for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[12] S. Amari, S. Douglas, A. Cichocki, and H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications*, Apr. 1997, pp. 101–104.

[13] V. Kautský, Z. Koldovský, and P. Tichavský, "Cramér-Rao-induced bound for interference-to-signal ratio achievable through non-gaussian independent component extraction," in *2017 IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2017, pp. 94–97.

[14] A. Hyvärinen, "One-unit contrast functions for independent component analysis: a statistical analysis," in *Neural Networks for Signal Processing VII. Proceedings of the 1997 IEEE Signal Processing Society Workshop*, Sep 1997, pp. 388–397.

[15] D.-T. A. Pham, "Blind partial separation of instantaneous mixtures of sources," in *Proceedings of International Conference on Independent Component Analysis and Signal Separation*. Springer Berlin Heidelberg, 2006, pp. 868–875.

[16] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge University Press, 2010.

[17] T. Menni, E. Chaumette, P. Larzabal, and J. P. Barbot, "New results on deterministic Cramér–Rao bounds for real and complex parameters," in *IEEE Trans. Signal Processing*, vol. 60, March 2012, pp. 1032–1049.

[18] B. Loesch and B. Yang, "Cramér–Rao bound for circular and noncircular complex independent component analysis," in *IEEE Trans. Signal Processing*, vol. 61, Jan 2013, pp. 365–379.

[19] J. Goodman and J. Weare, "Ensemble samplers with affine invariance," in *Communications in Applied Mathematics and Computational Science*, vol. 5, 2010, pp. 65–80.