CRAMÉR-RAO-INDUCED BOUND FOR INTERFERENCE-TO-SIGNAL RATIO ACHIEVABLE THROUGH NON-GAUSSIAN INDEPENDENT COMPONENT EXTRACTION

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 Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec,
Studentská 2, 461 17 Liberec, Czech Republic
[3]Institute of Information Theory and Automation, P.O.Box 18, 182 08 Prague 8, Czech Republic

ABSTRACT

This paper deals with the Cramér-Rao Lower Bound (CRLB) for a novel blind source separation method called Independent Component Extraction (ICE). Compared to Independent Component Analysis (ICA), ICE aims to extract only one independent signal from a linear mixture. The target signal is assumed to be non-Gaussian, while the other signals, which are not separated, are modeled as a Gaussian mixture. A CRLB-induced Bound (CRIB) for Interference-to-Signal Ratio (ISR) is derived. Numerical simulations compare the CRIB with the performance of an ICA and an ICE algorithm. The results show good agreement between the theory and the empirical results.

1. INTRODUCTION

In the fundamentals of Independent Component Analysis (ICA), the instantaneous linear mixing model

$$\mathbf{x} = \mathbf{A}\mathbf{u} \tag{1}$$

is studied [1,2]. Here, $\mathbf{x}$ is a $d\times 1$ vector of d mixed signals, $\mathbf{A}$ is a $d\times d$ non-singular mixing matrix, and $\mathbf{u}$ is a $d\times 1$ vector of the original signals that are assumed to be *mutually independent*. The jth signal u_j (the jth element of $\mathbf{u}$) is modeled as a random variable with the probability density function (pdf) $p_j(\cdot)$. The goal is to estimate $\mathbf{A}^{-1}$ from $\mathbf{x}$ through finding a square de-mixing matrix $\mathbf{W}$ such that $\mathbf{y} = \mathbf{W}\mathbf{x}$ are as independent as possible. In this paper, we will assume real-valued signals and parameters.

Many algorithms to solve this problem have been developed; see, e.g., [3]. It is known that if, at most, one original signal has Gaussian pdf while the other signals are non-Gaussian, then $\mathbf{A}^{-1}$ can be identified up to the order and scales of its rows [4]. It means that the de-mixing matrix $\mathbf{W}$ can be estimated as such that $\mathbf{G} = \mathbf{W}\mathbf{A} \approx \mathbf{P}\boldsymbol{\Lambda}$, where $\mathbf{P}$ and $\boldsymbol{\Lambda}$ is, respectively, a permutation and a diagonal matrix. The elements of $\mathbf{G}$ determine the accuracy of the separation. Its ijth element, G_{ij}, determines the presence of u_j in the ith separated signal y_i. The Cramér-Rao Lower Bound (CRLB) on the variance of G_{ij} provides an algorithm-independent bound for the estimation accuracy (for unbiased estimators). Using the CRLB theory, it is known, for the non-Gaussian ICA, that

$$\mathrm{E}[G_{ij}^2] \geq \frac{1}{N}\frac{\kappa_j}{\kappa_i\kappa_j - 1}, \qquad i \neq j, \tag{2}$$

where $\mathrm{E}[\cdot]$ stands for the expectation operator, N is the number of samples of $\mathbf{x}$ (assuming identically and independently distributed samples), and $\kappa_i = \mathrm{E}[\psi_i^2]$ where $\psi_i(x) = -\partial/\partial x \log p_i(x)$, which is called the score function of p_i where p_i is the pdf of the ith signal. For normalized variables with unit variance it holds that $\kappa_i \geq 1$ where $\kappa_i = 1$ if and only if the ith pdf is Gaussian; see, e.g., [5,6].

Recently, we have introduced a novel approach called Independent Component Extraction (ICE) in [7]. Here, the goal is to separate only one independent signal from $\mathbf{x}$ using a priori knowledge such as an initial guess (to determine which signal should be extracted). Without any loss of generality, let the signal of interest be $s = u_1$. In ICE, the mixing model (1) is re-parameterized for the extraction of s in the way that the rest of the mixture is not object of any particular decomposition, as compared to ICA. The motivation behind is that ICE algorithms could solve the simpler problem (to extract only one signal) faster that ICA methods; their complexities grow linearly with d while the complexities of ICA methods grow, at least, quadratically.

In this paper, we derive a lower bound on the achievable separation accuracy by ICE and compare it with (2). We consider a particular statistical model of signals. The signal of interest s is assumed to be non-Gaussian while the rest of the mixture is modeled as Gaussian. The latter is motivated by the fact that the other signals are never separated from each other (up to very special cases), so their joint distribution is close to

This work was supported by The Czech Science Foundation through Project No. 17-00902S.

Gaussian even if the pdfs of $u_2, \ldots, u_d$ are non-Gaussian.

The ICE mixing model parametrization and the statistical model of signals are described in Section 2. Section 3 is devoted to the computation of the Fisher Information Matrix, which is used in Section 4 to derive the Cramér-Rao Induced Bound (CRIB) for the achievable Interference-to-Signal Ratio (ISR). Section 5 is devoted to simulations and comparisons. Conclusions are drawn in Section 6.

The following notation will be used throughout the article. Plain letters denote scalars, bold letters denote vectors, and bold capital letters denote matrices. The Matlab convention for matrix/vector concatenation and indexing will be used, e.g., $[1;\ \mathbf{g}] = [1,\ \mathbf{g}^T]^T$, and $(\mathbf{A})_{j,:}$ is the jth row of $\mathbf{A}$.

2. PROBLEM STATEMENT

2.1. Mixing Model

ICE is based on a "model of model" as it is derived through a re-parameterization of (1). Let the mixing matrix $\mathbf{A}$ be partitioned as $\mathbf{A} = [\mathbf{a},\ \mathbf{A}_2]$, and let $\mathbf{x}$ be written as $\mathbf{x} = \mathbf{A}\mathbf{u} = \mathbf{a}s + \mathbf{y}$, where $\mathbf{y} = \mathbf{A}_2\mathbf{u}_2$ and $\mathbf{u}_2 = [u_2, \ldots, u_d]^T$. Since, in ICE, neither the identification of $\mathbf{A}_2$ nor the decomposition of $\mathbf{y}$ into independent signals is needed, the mixing matrix is considered as $\mathbf{A}_{\mathrm{ICE}} = [\mathbf{a},\ \mathbf{Q}]$ where $\mathbf{Q}$ is, for now, arbitrary, and the mixture is written as

$$\mathbf{x} = \mathbf{A}_{\mathrm{ICE}}\mathbf{v}, \tag{3}$$

where $\mathbf{v} = [s; \mathbf{z}]$, and $\mathbf{y} = \mathbf{Q}\mathbf{z}$. It holds that $\mathbf{z}$ belongs to the same subspace as $\mathbf{u}_2$ but is not explicitly determined yet.

The latter formulation, in fact, corresponds to a special case of Multidimensional ICA [8] (MICA) where the goal is to find two independent subspaces of dimensions 1 and $d-1$, namely, s and $\mathbf{z}$. However, ICE reduces the ambiguity of MICA through the following parameterization of the de-mixing matrix, denoted as $\mathbf{W}_{\mathrm{ICE}}$.

Let $\mathbf{a}$ and $\mathbf{W}_{\mathrm{ICE}}$ be partitioned, respectively, as $\mathbf{a} = [\gamma;\ \mathbf{g}]$ and $\mathbf{W}_{\mathrm{ICE}} = [\mathbf{w}^T;\ \mathbf{B}]$. $\mathbf{B}$ is required to be orthogonal to $\mathbf{a}$, which ensures that the signals separated by the lower part of $\mathbf{W}_{\mathrm{ICE}}$, that is $\mathbf{B}\mathbf{x}$, do not contain any contribution by s. A straightforward selection is $\mathbf{B} = [\mathbf{g} \quad -\gamma\mathbf{I}_{d-1}]$ where $\mathbf{I}_d$ denotes the $d \times d$ identity matrix. The free variables of $\mathbf{W}_{\mathrm{ICE}}$ are therefore represented by the elements of $\mathbf{a}$ and $\mathbf{w}$; let $\mathbf{w} = [\beta;\ \mathbf{h}]$. Hence

$$\mathbf{W}_{\mathrm{ICE}} = \begin{pmatrix} \mathbf{w}^T \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \beta & \mathbf{h}^T \\ \mathbf{g} & -\gamma\mathbf{I}_{d-1} \end{pmatrix}. \tag{4}$$

The next condition is that $\mathbf{W}_{\mathrm{ICE}}$ should be the inverse matrix of $\mathbf{A}_{\mathrm{ICE}}$, which guarantees that $s = \mathbf{w}^T\mathbf{x}$. This way $\mathbf{Q}$ and $\mathbf{z}$ can be determined. The reader can verify that the choice

$$\mathbf{A}_{\mathrm{ICE}} = [\mathbf{a}, \quad \mathbf{Q}] = \begin{pmatrix} \gamma & \mathbf{h}^T \\ \mathbf{g} & \frac{1}{\gamma}\left(\mathbf{g}\mathbf{h}^T - \mathbf{I}_{d-1}\right) \end{pmatrix}, \tag{5}$$

where β and γ are constrained to satisfy $\beta\gamma = 1 - \mathbf{h}^T\mathbf{g}$, guarantees that $\mathbf{W}_{\mathrm{ICE}}\mathbf{A}_{\mathrm{ICE}} = \mathbf{I}_d$.

The scales of s and of $\mathbf{a}$ are ambiguous in the sense that they can be substituted, respectively, by αs and $\alpha^{-1}\mathbf{a}$ where α is arbitrary such that $\alpha \neq 0$. This ambiguity can be removed by fixing β or γ. It is practical to select $\gamma = 1$ as in [7], because then the scale of s corresponds to the image of that source on the first sensor.

By adopting the idea of ICA, that is, taking the assumption that s and $\mathbf{z}$ are *independent*, ICE can be formulated as follows: *Find vectors* $\mathbf{g}$ *and* $\mathbf{h}$ *such that* $\mathbf{w}^T\mathbf{x}$ *and* $\mathbf{B}\mathbf{x}$*, where* $\mathbf{w} = [1 - \mathbf{h}^T\mathbf{g};\ \mathbf{h}]$ *and* $\mathbf{B} = [\mathbf{g}, \quad -\mathbf{I}_{d-1}]$*, are independent (or as independent as possible).*

2.2. Statistical Model

Several stochastic models have been considered in BSS/ICA to model the signals' independence, e.g., relying on one or more signal properties such as non-Gaussianity, nonstationarity or nonwhiteness [9]. The non-Gaussian ICA model (1) where all (but one) signals are non-Gaussian i.i.d. sequences is the most popular one. In ICE, there are only two variables: s, which coincides with u_1, and $\mathbf{z}$, which is a vector variable having an unspecified structure (it is a mixture of $u_2, \ldots, u_d$).

As in [7], we will assume that (1) s has a non-Gaussian pdf denoted as $p(s)$, while (2) $\mathbf{z}$ has multivariate Gaussian pdf with covariance $\mathbf{C}_{\mathbf{z}}$. The latter assumption can be justified by the fact that, up to very special cases, $\mathbf{z}$ is a mixture of $\mathbf{u}_2$. Even if $u_2, \ldots, u_d$ are non-Gaussian, their mixture tends to have distribution close to Gaussian due to the Central Limit Theorem [6]. The ICA and the ICE models coincide when $u_2, \ldots, u_d$ are Gaussian.

Hence, from (3), the pdf of $\mathbf{x}$ is

$$p_{\mathbf{x}}(\mathbf{x}) = p_s(\mathbf{w}^T\mathbf{x})p_{\mathbf{z}}(\mathbf{B}\mathbf{x})|\det(\mathbf{W}_{\mathrm{ICE}})|, \tag{6}$$

where $\mathbf{W}_{\mathrm{ICE}}$, $\mathbf{w}$, and $\mathbf{B}$ depend on $\mathbf{g}$ and $\mathbf{h}$ as described by (4), and $p_{\mathbf{z}}$ corresponds to $\mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{z}})$. A straightforward calculus, not shown here to save space, can show that for $\gamma = 1$, $|\det(\mathbf{W}_{\mathrm{ICE}})| = 1$. Hence, the log-likelihood function for one signal sample is equal to

$$\mathcal{L}(\mathbf{x}|\boldsymbol{\xi}) = \log p_s(\mathbf{w}^T\mathbf{x}) - \frac{1}{2}\mathbf{x}^T\mathbf{B}^T\mathbf{C}_{\mathbf{z}}^{-1}\mathbf{B}\mathbf{x} - \frac{1}{2}\log\left(|\mathbf{C}_{\mathbf{z}}|\right) - (d-1)\log\sqrt{2\pi}, \tag{7}$$

where $\boldsymbol{\xi} = [\mathbf{g}; \mathbf{h}; \mathbf{c}]$ is the parameter vector, in which $\mathbf{c}$ denotes the $d(d-1)/2 \times 1$ vector stacking the elements of $\mathbf{C}_z^{-1}$; $|\mathbf{C}_{\mathbf{z}}|$ denotes the determinant of $\mathbf{C}_{\mathbf{z}}$. Note that $\mathbf{c}$ is a nuisance parameter whose value cannot be treated as being known.

3. FISHER INFORMATION MATRIX

The Fisher information matrix $\mathbf{F}$ is defined as [10]

$$\mathbf{F}(\boldsymbol{\xi}) = \mathbb{E}\left[\frac{\partial\mathcal{L}}{\partial\boldsymbol{\xi}}\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\xi}}\right)^T\right]. \tag{8}$$

Let $\mathbf{F}$ be partitioned as

$$\mathbf{F}(\mathbf{g},\mathbf{h},\mathbf{c}) = \begin{pmatrix} \mathbf{F}_{\mathbf{g},\mathbf{g}} & \mathbf{F}_{\mathbf{g},\mathbf{h}} & \mathbf{F}_{\mathbf{g},\mathbf{c}} \\ \mathbf{F}_{\mathbf{h},\mathbf{g}} & \mathbf{F}_{\mathbf{h},\mathbf{h}} & \mathbf{F}_{\mathbf{h},\mathbf{c}} \\ \mathbf{F}_{\mathbf{c},\mathbf{g}} & \mathbf{F}_{\mathbf{c},\mathbf{h}} & \mathbf{F}_{\mathbf{c},\mathbf{c}} \end{pmatrix}. \tag{9}$$

The derivatives of the log-likelihood function (7) are:

$$\frac{\partial\mathcal{L}(\mathbf{x}|\mathbf{g},\mathbf{h},\mathbf{c})}{\partial\mathbf{g}} = \psi(s)\mathbf{h}x_1 - x_1\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}, \tag{10}$$

$$\frac{\partial\mathcal{L}(\mathbf{x}|\mathbf{g},\mathbf{h},\mathbf{c})}{\partial\mathbf{h}} = \psi(s)\mathbf{z}, \tag{11}$$

$$\frac{\partial\mathcal{L}(\mathbf{x}|\mathbf{g},\mathbf{h},\mathbf{c})}{\partial c_{i,j}} = -\frac{1}{2}\mathbf{z}^T\mathbf{J}^{i,j}\mathbf{z} + \frac{1}{2}\mathtt{tr}\left(\mathbf{J}^{i,j}\mathbf{C}_\mathbf{z}\right), \tag{12}$$

where $\mathbf{J}^{i,j}$ is the $d-1\times d-1$ matrix of zeros up to the ijth elements, which is equal to one; $\psi(\cdot)$ is the score function of $p(\cdot)$. The blocks of $\mathbf{F}$ are as follows:

$$\begin{aligned}\mathbf{F}_{\mathbf{g},\mathbf{g}} &= \mathbb{E}[\psi^2(s)x_1^2\mathbf{h}\mathbf{h}^T - \psi(s)x_1^2\mathbf{h}\mathbf{z}^T\mathbf{C}_\mathbf{z}^{-1} - \\ &\quad - \psi(s)x_1^2\mathbf{z}\mathbf{h}^T\mathbf{C}_\mathbf{z}^{-1} + x_1^2\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}\mathbf{z}^T\mathbf{C}_\mathbf{z}^{-1}] = \\ &= \mathbb{E}[\eta\mathbf{h}\mathbf{h}^T + \kappa y_1^2\mathbf{h}\mathbf{h}^T - 2\psi(s)sy_1\mathbf{h}\mathbf{z}^T\mathbf{C}_\mathbf{z}^{-1} - \\ &\quad - 2\psi(s)sy_1\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}\mathbf{h}^T + (s^2+y_1^2)\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}\mathbf{z}^T\mathbf{C}_\mathbf{z}^{-1}] = \\ &= \eta\mathbf{h}\mathbf{h}^T + \kappa(\mathbf{h}^T\mathbf{C}_\mathbf{z}\mathbf{h})\mathbf{h}\mathbf{h}^T - 4\mathbf{h}\mathbf{h}^T + \sigma_s^2\mathbf{C}_\mathbf{z}^{-1} + \\ &\quad + \mathbb{E}[\mathbf{C}_\mathbf{z}^{-1}(\mathbf{h}^T\mathbf{z})^2\mathbf{z}\mathbf{z}^T\mathbf{C}_\mathbf{z}^{-1}] = \\ &= \eta\mathbf{h}\mathbf{h}^T + \kappa(\mathbf{h}^T\mathbf{C}_\mathbf{z}\mathbf{h})\mathbf{h}\mathbf{h}^T - 2\mathbf{h}\mathbf{h}^T + \sigma_s^2\mathbf{C}_\mathbf{z}^{-1} + \\ &\quad + \mathbf{C}_\mathbf{z}^{-1}(\mathbf{h}^T\mathbf{C}_\mathbf{z}\mathbf{h}),\end{aligned} \tag{13}$$

where

$$\kappa = \mathbb{E}[\psi^2(s)], \tag{14}$$

$$\eta = \mathbb{E}[\psi^2(s)s^2], \tag{15}$$

$$\sigma_s^2 = \mathbb{E}[s^2], \tag{16}$$

and where the following identities have been used

$$\mathbb{E}[\psi(s)s] = 1, \tag{17}$$

$$y_1 = \mathbf{h}^T\mathbf{z}, \tag{18}$$

$$\mathbb{E}\left[y_1^2\mathbf{y}\mathbf{y}^T\right] = \mathbb{E}\left[y_1^2\right]\mathbb{E}\left[\mathbf{y}\mathbf{y}^T\right] + 2\mathbb{E}\left[y_1\mathbf{y}\right](\mathbb{E}\left[y_1\mathbf{y}\right])^T. \tag{19}$$

Next,

$$\mathbf{F}_{\mathbf{h},\mathbf{h}} = \mathbb{E}\left[\psi^2(s)\mathbf{z}\mathbf{z}^T\right] = \kappa\mathbf{C}_\mathbf{z}, \tag{20}$$

$$(\mathbf{F}_{\mathbf{c},\mathbf{c}})_{i,j} = \frac{1}{2}\left(\mathtt{tr}(\mathbf{J}^{i,j}\mathbf{C}_\mathbf{z}\mathbf{J}^{j,i}\mathbf{C}_\mathbf{z})\right), \tag{21}$$

$$\begin{aligned}\mathbf{F}_{\mathbf{g},\mathbf{h}} &= \mathbb{E}\left[\psi^2(s)x_1\mathbf{h}\mathbf{z}^T - x_1\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}\mathbf{z}^T\psi(s)\right] = \\ &= \kappa\mathbf{h}\mathbf{h}^T\mathbf{C}_\mathbf{z} - \mathbf{I}_{d-1},\end{aligned} \tag{22}$$

$$\mathbf{F}_{\mathbf{h},\mathbf{c}} = \mathbf{O}, \tag{23}$$

where $\mathbf{O}$ denotes a zero matrix of corresponding dimension. Finally,

$$\begin{aligned}(\mathbf{F}_{\mathbf{g},\mathbf{c}})_{:,k} &= \frac{1}{2}\mathbb{E}\left[-\psi(s)\mathbf{h}x_1\mathbf{z}^T\mathbf{J}^{i,j}\mathbf{z}\right] + \\ &\quad + \frac{1}{2}\mathbb{E}\left[x_1\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}\mathbf{z}^T\mathbf{J}^{i,j}\mathbf{z}\right] + \\ &\quad + \frac{1}{2}\mathbb{E}\left[\mathbf{h}\cdot\mathtt{tr}(\mathbf{J}^{j,i}\mathbf{C}_\mathbf{z}) - \mathbf{h}\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}\mathbf{z}^T\mathtt{tr}(\mathbf{J}^{j,i}\mathbf{C}_\mathbf{z})\right] = \\ &= -\frac{1}{2}\mathbf{h}\mathbb{E}\left[\mathbf{z}^T\mathbf{J}^{i,j}\mathbf{z}\right] + \frac{1}{2}\mathbb{E}\left[y_1\mathbf{C}_\mathbf{z}^{-1}\mathbf{z}\mathbf{z}^T\mathbf{J}^{i,j}\mathbf{z}\right] = \\ &= -\frac{1}{2}\mathbf{h}\cdot\mathtt{tr}(\mathbf{J}^{i,j}\mathbf{C}_\mathbf{z}) + \frac{1}{2}\mathbf{C}_\mathbf{z}^{-1}\mathbf{V}_{:,ij},\end{aligned} \tag{24}$$

where

$$\begin{aligned}\mathbf{V}_{:,ij} &= \mathbb{E}[y_1 z z_i z_j] = \mathbb{E}[y_1 z]\mathbb{E}[z_i z_j] + \mathbb{E}[y_1 z_i]\mathbb{E}[z z_j] + \\ &\quad + \mathbb{E}[y_1 z_j]\mathbb{E}[z z_i] = \\ &= \mathbf{C}_\mathbf{z}\mathbf{h}(\mathbf{C}_\mathbf{z})_{ij} + (\mathbf{C}_\mathbf{z})_{i,:}\mathbf{h}(\mathbf{C}_\mathbf{z})_{:,j} + (\mathbf{C}_\mathbf{z})_{j,:}\mathbf{h}(\mathbf{C}_\mathbf{z})_{:,i},\end{aligned} \tag{25}$$

$i = 1,\ldots,d-1$, $j = 1,\ldots,d-1$ and $i > j$. The other blocks follow the symmetry of FIM. Now, the CRLB is obtained through the computation of $\mathbf{F}^{-1}$.

4. CRLB-INDUCED BOUND FOR ISR

Here, we derive the lower bound for the achievable mean value of the Interference-to-Signal Ratio (ISR) using the CRLB derived above. Let $\widehat{\mathbf{w}}$ be an estimated vector that separates s from $\mathbf{x}$, and let $\mathbf{A}_{\mathrm{ICE}}$ be the true mixing matrix in (3). Then, the ISR of the extracted signal $\widehat{s} = \widehat{\mathbf{w}}^T\mathbf{x}$ is

$$\mathtt{ISR} = \frac{\mathbb{E}[(\widehat{\mathbf{w}}^T\mathbf{y})^2]}{\mathbb{E}[(\widehat{\mathbf{w}}^T\mathbf{a}s)^2]} = \frac{\mathbf{q}_2^T\mathbf{C}_\mathbf{z}\mathbf{q}_2}{q_1^2\sigma_s^2} \approx \frac{1}{\sigma_s^2}\mathbf{q}_2^T\mathbf{C}_\mathbf{z}\mathbf{q}_2, \tag{26}$$

where $\mathbf{q}^T = [q_1,\, \mathbf{q}_2^T] = \widehat{\mathbf{w}}^T\mathbf{A}_{\mathrm{ICE}} = [\widehat{\mathbf{w}}^T\mathbf{a},\, \widehat{\mathbf{w}}^T\mathbf{Q}]$. The last approximation assumes "small" errors, so $q_1^2 \approx 1$ and $\mathbf{q} \approx \mathbf{e}_1$ (the unit vector). Then, the mean ISR value reads

$$\mathbb{E}\left[\mathtt{ISR}\right] \approx \frac{1}{\sigma_s^2}\mathbb{E}\left[\mathbf{q}_2^T\mathbf{C}_\mathbf{z}\mathbf{q}_2\right] = \frac{1}{\sigma_s^2}\mathtt{tr}\left(\mathbf{C}_\mathbf{z}\mathbb{E}\left[\mathbf{q}_2\mathbf{q}_2^T\right]\right). \tag{27}$$

For further analysis, let us consider the special case when $\mathbf{C}_\mathbf{z} = \sigma_\mathbf{z}^2\mathbf{I}_{d-1}$. Then, (27) simplifies to

$$\mathbb{E}\left[\mathtt{ISR}\right] \approx \frac{\sigma_\mathbf{z}^2}{\sigma_s^2}\mathtt{tr}\left(\mathbb{E}\left[\mathbf{q}_2\mathbf{q}_2^T\right]\right) = \frac{\sigma_\mathbf{z}^2}{\sigma_s^2}\mathtt{tr}\left(\mathtt{cov}\left(\mathbf{q}_2\right)\right). \tag{28}$$

Thanks to the equivariance property [5], we can consider the special case when $\mathbf{g} = \mathbf{h} = \mathbf{0}$. Then, it holds that $\mathbf{q}_2 = \widehat{\mathbf{h}}$, where $\widehat{\mathbf{h}}$ is $\widehat{\mathbf{w}}$ without its first element, hence,

$$\mathbb{E}\left[\mathtt{ISR}\right] \approx \frac{\sigma_\mathbf{z}^2}{\sigma_s^2}\mathtt{tr}(\mathtt{cov}(\widehat{\mathbf{h}})) \geq \frac{\sigma_\mathbf{z}^2}{\sigma_s^2}\mathtt{tr}\left(\mathtt{CRLB}\left(\mathbf{h}\right)\right), \tag{29}$$

where $\mathtt{CRLB}(\mathbf{h})$ denotes the diagonal block of $\mathbf{F}(\boldsymbol{\xi})^{-1}$ corresponding to the parameter $\mathbf{h}$. Since $\mathbf{h} = \mathbf{g} = \mathbf{0}$, (8) is equal to

$$\mathbf{F}(\boldsymbol{\xi}) = \begin{pmatrix} \frac{\sigma_s^2}{\sigma_{\mathbf{z}}}\mathbf{I}_{d-1} & -\mathbf{I}_{d-1} & \mathbf{O} \\ -\mathbf{I}_{d-1} & \kappa\sigma_{\mathbf{z}}^2\mathbf{I}_{d-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{D} \end{pmatrix}, \tag{30}$$

where $\mathbf{D} = \frac{1}{2}\sigma_{\mathbf{z}}^4\mathbf{I}_{\frac{d(d-1)}{2}}$. Then, it holds that

$$\mathbf{F}(\boldsymbol{\xi})^{-1} = \begin{pmatrix} \frac{\kappa\sigma_{\mathbf{z}}^2}{\kappa\sigma_s^2-1}\mathbf{I}_{d-1} & \frac{1}{\kappa\sigma_s^2-1}\mathbf{I}_{d-1} & \mathbf{O} \\ \frac{1}{\kappa\sigma_s^2-1}\mathbf{I}_{d-1} & \frac{1}{\kappa\sigma_s^2-1}\frac{\sigma_s^2}{\sigma_{\mathbf{z}}^2}\mathbf{I}_{d-1} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{D^{-1}} \end{pmatrix}. \tag{31}$$

Using (29) and (31), and by considering N observations, the CRLB-induced bound for ISR is obtained as

$$\mathbb{E}\left[\mathtt{ISR}\right] \geq \frac{1}{N}\frac{d-1}{\kappa\sigma_s^2 - 1}\frac{\sigma_s^2}{\sigma_{\mathbf{z}}^2}. \tag{32}$$

To compare this bound with (2), let us consider $\sigma_s^2 = \sigma_{\mathbf{z}}^2 = 1$. Then, $\mathbb{E}\left[\mathtt{ISR}\right] \approx \sum_{k=2}^{d}\mathbb{E}\left[G_{1,k}^2\right]$ and (32) simplifies to

$$\mathbb{E}\left[\mathtt{ISR}\right] \geq \frac{d-1}{N}\frac{1}{\kappa - 1}. \tag{33}$$

Moreover, assume that $u_1, \ldots, u_d$ have all unit variance and that $u_2, \ldots, u_d$ have all Gaussian pdf, which means that $\kappa_j = 1$ for $j = 2, \ldots, d$. In that special case, it can be verified, using (2), that the induced bound for ICA coincides with that for ICE given by (33). Also, (33) coincides with the optimum theoretical performance of methods performing one-unit (or partial) source separation [5, 11, 12].

5. SIMULATIONS

In simulations, we compare the bound for ICE with empirical mean ISR achieved by two methods: EFICA (Efficient Fast ICA) and OGICE (Orthogonally Constrained ICE). EFICA is an ICA algorithm that approaches the CRLB when pdfs of all components belong to the Generalized Gaussian Distribution (GGD) family [13]. OGICE is a recently developed method for ICE that has to be properly initialized, i.e., within the area of convergence to the desired (target) signal [7]. OGICE is derived based on maximum likelihood principle, so it might achieve asymptotic efficiency, however, it exploits the orthogonal constraint (the subspaces of the target signal and of the other signals are constrained to be orthogonal), which can cause performance limitations [5].

In a trial, $d = 5$ independent signals are generated. The target signal is drawn from a GGD with the shape parameter $\alpha \in (0, +\infty)$. The other signals are Gaussian, which corresponds to $\alpha = 2$. All signals are normalized to have zero mean and unit variance and mixed by a random mixing matrix $\mathbf{A}$ as in (1). Note that such mixture corresponds with (3) where $\mathbf{C_z}$ is determined by $\mathbf{A}$. OGICE is initialized by a randomly perturbed first column of $\mathbf{A}$, while the initialization of EFICA is random in full (the separated target signal is determined based on ISR).

Figures 1 and 2 show average ISR achieved by the algorithms in 1000 trials, respectively, for varying N (when $\alpha = 1$) and varying α ($N = 2500$). The results from EFICA confirm the validity of the bound (33); the ISRs achieved by the algorithms are very close almost in all cases. However, OGICE does not converge to global maximum in each trial. Hence, we take the average ISR for this method only over trials where $\mathtt{ISR} \leq \frac{1}{10}$ (to remove those trials where the algorithm did not converge). Therefore, the resulting ISR by OGICE does not fully obey the bound (33) (can be sometimes lower, especially, for α close to 2).

With growing N, the ISRs are decreasing. For $\alpha = 2$, where all signals (including the target one) are Gaussian and are not separable, the ISRs approach 0 dB, which means poor separation accuracy.

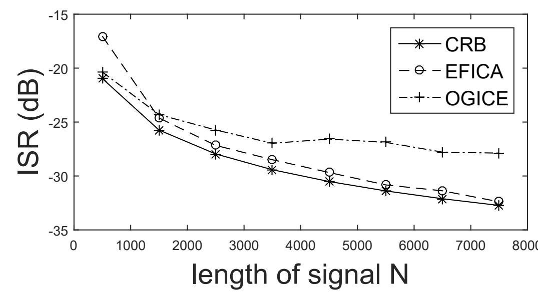

Fig. 1. Average ISR for $d = 5$, $\alpha = 1$, and varying N.

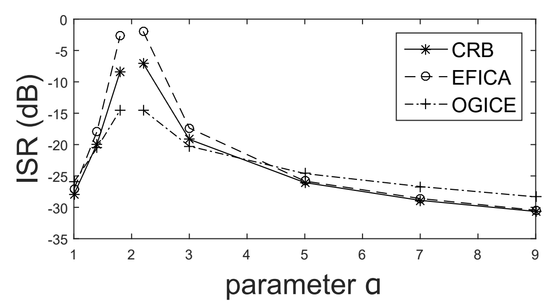

Fig. 2. Average ISR for $d = 5$, $N = 2500$ and varying α.

6. CONCLUSIONS

The CRIB for achievable ISR through ICE was shown to be attainable by EFICA, depending on the target signal pdf and the length of data. The bound coincides with that for ICA provided that the rest of the linear mixture is Gaussian. Future works will be focused on situations when the rest is a mixture of non-Gaussian signals. Then, the open question is whether other than Gaussian modeling enables ICE to approach the CRIB achievable through ICA.

7. REFERENCES

[1] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, pp. 287–314, 1994.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.

[3] P. Tichavský and Z. Koldovský, "Fast and accurate methods of independent component analysis: A survey," *Kybernetika*, vol. 47, no. 3, pp. 426–438, 2011.

[4] 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.

[5] P. Tichavský, Z. Koldovský, and E. Oja, "Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, pp. 1189–1203, April 2006.

[6] J. F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct 1998.

[7] 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.

[8] J. F. Cardoso, "Multidimensional independent component analysis," in *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, vol. 4, May 1998, pp. 1941–1944 vol.4.

[9] T. Adalı, M. Anderson, and G. S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, May 2014.

[10] A. Papoulis and S. U. Pillai, *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 2002.

[11] 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.

[12] D.-T. A. Pham, *Blind Partial Separation of Instantaneous Mixtures of Sources*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 868–875.

[13] Z. Koldovský, P. Tichavský, and E. Oja, "Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1265–1277, Sept 2006.