Cramér-Rao Bounds for Complex-Valued Independent Component Extraction: Determined and Piecewise Determined Mixing Models

Václav Kautský, Zbyněk Koldovský, Petr Tichavský, Vicente Zarzoso

Authors' manuscript
Accepted for publication in *IEEE Transactions on Signal Processing*

Cramér-Rao Bounds for Complex-Valued Independent Component Extraction: Determined and Piecewise Determined Mixing Models

Václav Kautský[1,2], Zbyněk Koldovský[2], Petr Tichavský[3], and Vicente Zarzoso[4], *Senior Member, IEEE*

***Abstract*—Blind source extraction (BSE) aims at recovering an unknown source signal of interest from the observation of instantaneous linear mixtures of the sources. This paper presents Cramér-Rao Lower Bound (CRLB) for the complex-valued BSE problem based on the assumption that the target signal is independent of the other signals. The target source is assumed to be non-Gaussian or non-circular Gaussian while the other signals (background) are circular Gaussian or non-Gaussian. The results confirm some previous observations known for the real domain and yield new results for the complex domain. Also, the CRLB for Independent Component Extraction (ICE) is shown to coincide with that for Independent Component Analysis (ICA) when the non-Gaussianity of background is taken into account. Second, we extend the CRLB analysis to piecewise determined mixing models, where the observed signals are assumed to obey the determined mixing model within short blocks where the mixing matrices can be varying from block to block. This model has applications, for instance, when separating dynamic mixtures. Either the mixing vector or the separating vector corresponding to the target source is assumed to be constant across the blocks. The CRLBs for the parameters of these models bring new performance limits for the BSE problem.**

I. Introduction

Blind source separation (BSS) aims at recovering a set of unobservable signals, called sources, from a set of observed mixtures of the sources [1]. This problem has drawn a lot of attention from the signal processing and machine learning communities over the last two decades, especially due to the vast amount of application domains where it is pertinent and has produced useful results. When the sources are statistically independent, BSS can be solved through the statistical tool of independent component analysis (ICA). Blind source extraction (BSE) is a related problem where the goal is to estimate a particular source of interest (SOI) in the set of unobservable signals. BSE is motivated by the fact that, often, targeting the SOI may be considerably more cost-effective than separating the whole set of sources from the observed mixture.

A wide variety of signal processing methods for BSS and BSE have been proposed in the literature; a thorough review can be found in [1]. Because different methods may typically provide different results, a fundamental question is the performance limits that can be attained in a given scenario regardless of the methods employed. Cramér-Rao lower bounds (CRLB) are useful for this task, and have been therefore studied, e.g., in [2]–[5].

The present paper focuses on the BSE problem where the SOI is assumed to be independent from the background, a problem closely related to ICA. Our contribution is two-fold. In the first place, the standard determined mixing scenario is considered, where the BSE problem is formulated through the recently proposed approach called independent component extraction (ICE) [6], based on a particular parameterization of the mixing system. In the second place, we focus on the *piecewise* determined mixing model that is usually designed for dynamic mixtures, e.g., the moving source in a static background is studied in [7].

In the assumed model, the observed samples are partitioned into several blocks where the samples in each block obey the standard determined model. Piecewise models extend the standard BSS/BSE problem, typically defined in the context of instantaneous mixtures, to the more general case of convolutive mixtures, as they naturally arise when transforming the observations into the frequency domain. We compute the CRLBs of two piecewise determined mixing models used for BSE.

The paper is organized as follows. The BSS and BSE problems are recalled in Section II, together with the main results found in the literature. Section III is devoted to the standard determined mixing model and the above mentioned issues related to the CRLBs. Section IV introduces the piecewise determined mixing models, and derives the related CRLBs using the results of Section III. The computed theoretical bounds are discussed and compared in Section V by analyzing several special cases. An experimental validation is presented in Section VI, while the conclusions of Section VII bring the paper to the end.

II. Problem Statement

A. Mathematical Notation

Throughout the paper, plain, bold lowercase and bold capital letters denote, respectively, scalars, vectors and matrices. Symbols $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^*$ denote, respectively, transposition, conjugate transpose and complex conjugate. The Matlab convention for matrix/vector concatenation and

[0]This work was supported by The Czech Science Foundation through Project No. 17-00902S, by the United States Department of the Navy, Office of Naval Research Global, through Projects No. N62909-18-1-2040 and N62909-19-1-2105 and by the grant SGS18/188/OHK4/3T/14.
[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] The Czech Academy of Sciences, Institute of Information Theory and Automation, P. O. Box 18, 182 08 Prague 8, Czech Republic
[4]Université Côte d'Azur, CNRS, I3S Laboratory, Sophia Antipolis, France.

indexing will be used, e.g., $[1;\,\mathbf{g}] = [1,\,\mathbf{g}^T]^T$, and $(\mathbf{A})_{j,:}$ is the jth row of $\mathbf{A}$. Notation $\mathrm{E}[\cdot]$ stands for the expectation operator. In this paper, complex-valued signals and parameters will be considered. A complex random vector $\mathbf{x}$ is called circular if its pseudo-covariance, defined as $\mathtt{pcov}(\mathbf{x}) = \mathrm{E}\big[(\mathbf{x}-\mathrm{E}[\mathbf{x}])\,(\mathbf{x}-\mathrm{E}[\mathbf{x}])^T\big]$, is null. Otherwise, $\mathbf{x}$ is non-circular. The second-order circularity coefficient ρ of a complex-valued random variable x with zero mean, see [8], is defined as in [9] $\rho = \big|\mathrm{E}\big[x^2\big]\big| \,/ \mathrm{E}\big[|x|^2\big]$. Thus, $\rho \in [0,1]$, and $\rho = 0$ for circular random variables.

B. Signal Models

Classical BSS considers the instantaneous linear mixing model

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

where $\mathbf{x}$ is a $d\times 1$ vector representing d observed signals, $\mathbf{u}$ is a $n\times 1$ vector of source signals, and $\mathbf{A}$ is a $d\times n$ mixing matrix. Throughout the paper, we will consider the more general case of complex-valued sources and mixing matrices; the real-valued case will be addressed in Section IV.F.

The goal of BSS is to separate $\mathbf{u}$ from $\mathbf{x}$ using only information provided by the observed samples [1]. BSE aims at separating only one source, referred to as the SOI, from the remaining sources in $\mathbf{x}$, which are called *background*. The standard model considers the so-called *determined* or *square* case, where the number of sources is the same as that of the observed signals, $n = d$ and $\mathbf{A}$ is hence a square $d \times d$ non-singular matrix. ICA is a popular BSS method based on the assumption that the source signals are mutually independent. Under this assumption, the estimation of $\mathbf{A}$ and of $\mathbf{A}^{-1}$ is equivalent to the separation of $\mathbf{u}$, which can be carried out by finding a square de-mixing matrix $\mathbf{W}$ such that

$$\widehat{\mathbf{u}} = \mathbf{W}\mathbf{x} \tag{2}$$

are as independent as possible. Identifiability and separability conditions are analyzed in [8].

Another interesting model often arises when separating convolutive mixtures in the frequency domain [10], in the problem of independent vector analysis (IVA) [11] or yet in joint BSS, where several instantaneous mixtures are observed:

$$\mathbf{x}^k = \mathbf{A}^k\mathbf{u}^k, \qquad k = 1,\ldots,K. \tag{3}$$

Here, k plays the role of the mixture or dataset index, e.g., the frequency bin index when transforming a convolutive mixture into the frequency domain. The source signals $\mathbf{u}^k = [u_1^k,\ldots,u_d^k]^T$ are assumed to be mutually independent while vector components $\mathbf{u}_i = [u_i^1,\ldots,u_i^K]^T$, $i = 1,\ldots,d$, have elements that can be mutually dependent. The latter property is exploited for joint separation of the set of mixtures.

In piecewise determined mixing models, it is assumed that the observed samples of mixed signals can be partitioned into M blocks where the samples in each block obey the standard determined model (1). The mth block is thus described by

$$\mathbf{x}^m = \mathbf{A}^m\mathbf{u}^m, \qquad m = 1,\ldots,M, \tag{4}$$

where the source signals $\mathbf{u}^m = [u_1^m,\ldots,u_d^m]^T$ are independent. The mixing matrices $\mathbf{A}^1,\ldots,\mathbf{A}^M$ as well as the source signals (their distributions) may vary from block to block. The model thus involves dynamic mixing as well as a special underdetermined case (more sources than sensors) since there can be up to Md sources. As we will see later in the paper, the fact that the mixtures are determined within the blocks allows the analytic computation of the CRLB.

The ideas of the joint mixing and of the piecewise determined mixing models can be yet combined together (dataset and block indices are needed) [12]. Also, since the algebraic definitions (3) and (4) are formally identical, IVA can be considered for solving the latter problem; see, e.g., [13]. In this paper, we will focus on the single-dataset mixing models (1) and (4); the other variants exceed the scope of this paper. Before presenting our contribution, we turn to a review of the existing literature.

C. Overview of Existing Results

1) Independence-based BSS/BSE methods: BSE methods based on source non-Gaussianity had been studied even before ICA was formulated [14], [15] in Comon's pioneering paper [16]. Then, the theory of ICA has been established since 90s; see, e.g., [1], [2], [17], [18]. The relation of the non-Gaussianity based BSE methods has been described through information theory and the properties of the Kullback-Leibler divergence (mutual information) and entropy [19]. ICE is a recent revision of this relation based on an algebraic mixing model, as will be recalled in Section III, and maximum likelihood estimation [6].

ICA has been used for blind separation of convolutive mixtures in the frequency domain [10], where the mixture is transformed into a set of complex-valued instantaneous mixtures, with one mixture per frequency. The problem, called frequency-domain ICA (FDICA), is formally described by (4), however, m plays the role of the frequency bin index. When ICA is applied separately to each mixture, the indeterminacy of the order of separated component gives rise to the permutation problem [20], and the separated frequency components must be reordered in order to allow for the separation of signals in the time domain.

To avoid the permutation problem, IVA has been proposed [11]. Here, the algebraic model given by (3) remains the same as in FDICA while the statistical model involves the assumption that independent components belonging to the same source are mutually dependent and form so-called vector components. The idea of IVA have become very popular due to its wide applicability far beyond audio source separation [21]–[23]. Its variant for BSE appeared, e.g., in [24], and has been recently formulated as independent vector extraction (IVE) in [6].

Another recent advance in this line represents independent low rank matrix analysis (ILRMA) where the statistical model of a vector component (representing one source) assumes that its spectrogram has a low-rank structure. For example, ILRMA combines IVA and nonnegative matrix factorization (NMF) in [25], [26].

In BSS/BSE, there is a wide class of methods based on Gaussian statistical signal models, as compared to the non-Gaussianity-based methods considered in this paper. Those

methods exploit only second order statistics (SOS) and their algebraic properties. For example, the analogy of the standard ICA problem based on SOS boils down to the problem of joint approximate diagonalization (JAD) of covariance matrices; see, e.g., [27]–[30] and references therein. Similarly to IVA, the SOS-based methods were considered in [31], [32]; see also [33].

2) Locally Determined Models for Underdetermined BSS: When the mixing model (1) involves more sources than observations ($n > d$), the source extraction/separation and the mixing matrix identification problems are no longer equivalent. Therefore, they are typically treated separately in two step procedures. For example, the estimation of $\mathbf{A}$ can be done by applying a decomposition to a tensor that is built from covariance matrices [34] or higher-order based statistics [35], [36]. Then, various array processing methods can be applied to extract the sources [37], [38].

There are also BSS methods that treat the underdetermined problem by assuming a certain local condition guaranteeing that the every sample or time-frequency point involves maximally d sources. Most typically, blind speech separation methods exploit the time-frequency sparsity of speech signals [39], [40]. Other methods assume that there are single-source points or regions and the separation mainly relies on a detection of these regions [41], [42]. Locally determined mixing is considered, e.g., in [43]. The ICA models presented in Section III could be considered as members of the class of locally determined models for BSE, where identification and extraction proceed jointly.

3) Performance bounds: Performance limitations of ICA based on the standard determined mixing model have been well investigated in the literature. It is known that $\mathbf{A}$ in (1) can be identified up to the order and scales of its columns if it holds that at most one source signal has the complex Gaussian pdf or that no two complex Gaussian source signals have the same circularity coefficient [8]. Then, a de-mixing matrix $\mathbf{W}$ can be estimated as such that $\mathbf{G} = \mathbf{WA} \approx \mathbf{P\Lambda}$, where $\mathbf{P}$ and $\mathbf{\Lambda}$ is a permutation and diagonal matrix (with nonzero diagonal entries), respectively. $\mathbf{G}$ reflects the separation accuracy as its ijth element, G_{ij}, determines the presence of u_j in the ith separated signal $\widehat{u}_i$, so there is a clear correspondence between the elements of $\mathbf{G}$ and the Interference-to-Signal Ratio (ISR) of the separated signals. For the real-valued (and similarly for the complex-valued) ICA problem, it was derived using the CRLB that the ISR of the ith separated source obeys

$$\mathrm{E}[\mathtt{ISR}_i] \geq \frac{1}{N} \sum_{j=1, j\neq i}^{d} \frac{\overline{\kappa}_j}{\overline{\kappa}_i \overline{\kappa}_j - 1} \tag{5}$$

where N is the number of i.i.d. samples [3], [5]; $\kappa_i = \mathrm{E}[|\psi_i|^2]$ where $\psi_i(x) = -\partial/\partial x \, \log p_i(x)$ is the score function related to p_i, and $\overline{\kappa}_i = \kappa_i \sigma_i^2$ where σ_i^2 is the variance of u_i; $\overline{\kappa}_i$ corresponds to κ_i when p_i is normalized to unit variance. It holds that $\overline{\kappa}_i \geq 1$, and $\overline{\kappa}_i = 1$ if and only if the ith pdf is circular Gaussian. Hence, the denominator in (5) approaches zero when both the ith and the jth source signals are close to circular Gaussian.

This brings some issues into question regarding the BSE problem. Without loss on generality, let $(d-1)$ source signals in the mixture be circular Gaussian but not so the first source (SOI). Then, $\mathbf{A}$ is no more identifiable, and the CRLB (5) formally does not exist. However, BSE methods exploiting the non-Gaussianity of the SOI are known for their ability to blindly extract that source; see, e.g., [3]. Moreover, their asymptotic performance analyses have shown that their accuracy is limited by

$$\mathrm{E}[\mathtt{ISR}] \geq \frac{1}{N} \frac{d-1}{\overline{\kappa} - 1}, \qquad i = 2, \ldots, d \tag{6}$$

where $\overline{\kappa} = \overline{\kappa}_1$; see, e.g., [3], [44], [45]. This asymptotic bound coincides with the right-hand side of (5) when considering $i = 1$ and $\overline{\kappa}_j = 1$ for $j = 2, \ldots, d$.

A formal confirmation of this bound for the real-valued case has been proven recently in [46] through computing the CRLB for the ICE mixing model, that is, assuming that the mixing matrix is structured as described by (8) and that the background signals are Gaussian.

D. Summary of our Contribution

In the first part of this paper (Section III), we generalize the above result for the complex-valued case where the SOI is assumed to be non-Gaussian or non-circular Gaussian. The background is modeled as circular Gaussian or circular non-Gaussian. We avoid the case with non-circular background, for simplicity, as it is computationally less tractable and its analysis goes beyond the scope of this paper. We show that the CRLB of ICE coincides with the bound for ICA when the background is circular Gaussian, as in the real-valued case. Moreover, we also show that these bounds coincide when the background modeling in ICE takes into account possible non-Gaussianity of the background. In the second part of the paper (Section IV), these results are generalized to the piecewise model (4) and to its special variants by extending the ICE parameterization. The results of the former part fill in the gaps currently existing in the theory of ICA/ICE performance bounds. To the best of our knowledge, the results of the latter part are completely original as this is the first work that considers the performance bounds of the piecewise determined mixtures.

III. CRLB for Determined Mixing

A. Algebraic Model

We begin our development by briefly explaining the ICE parameterization recently proposed in [6]. It is assumed, without loss of generality, that the separation system is designed to extract the first source u_1, which plays the role of the SOI in (1). Then the mixing matrix can be partitioned as $\mathbf{A} = [\mathbf{a}, \, \mathbf{A}_2]$ and the observations $\mathbf{x}$ can be written as

$$\mathbf{x} = \mathbf{Au} = \mathbf{a}s + \mathbf{y} \tag{7}$$

where $s = u_1$, $\mathbf{y} = \mathbf{A}_2\mathbf{u}_2$ and $\mathbf{u}_2 = [u_2, \ldots, u_d]^T$. Since neither $\mathbf{u}_2$ nor $\mathbf{A}_2$ need to be estimated in order to extract s, we can consider any auxiliary background signals $\mathbf{z}$ such that $\mathbf{y} = \mathbf{A}_2\mathbf{u}_2 = \mathbf{Qz}$, where the columns of $\mathbf{Q}$ span the

same subspace as that of the columns of $\mathbf{A}_2$. Compared to $\mathbf{u}_2$, the elements of $\mathbf{z}$ need not be independent, so $\mathbf{Q}$ can be arbitrary in this sense. Now, according to the ICE model, the mixing matrix and its inverse (de-mixing) matrix can be parameterized, respectively, as

$$\mathbf{A}_{\mathrm{ICE}} = \begin{pmatrix} \mathbf{a} & \mathbf{Q} \end{pmatrix} = \begin{pmatrix} \gamma & \mathbf{h}^H \\ \mathbf{g} & \frac{1}{\gamma}\left(\mathbf{g}\mathbf{h}^H - \mathbf{I}_{d-1}\right) \end{pmatrix} \tag{8}$$

$$\mathbf{W}_{\mathrm{ICE}} = \mathbf{A}_{\mathrm{ICE}}^{-1} = \begin{pmatrix} \mathbf{w}^H \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \beta^* & \mathbf{h}^H \\ \mathbf{g} & -\gamma\mathbf{I}_{d-1} \end{pmatrix} \tag{9}$$

where $\mathbf{a}$ denotes the first column of $\mathbf{A}$, which is the mixing vector related to u_1 partitioned as $\mathbf{a} = [\gamma; \mathbf{g}]$, and $\mathbf{w}$ is the separating vector such that $\mathbf{w}^H\mathbf{x} = u_1$, partitioned as $\mathbf{w} = [\beta; \mathbf{h}]$. Symbol $\mathbf{I}_d$ denotes the $d \times d$ identity matrix, and β and γ are linked through

$$\beta^*\gamma = 1 - \mathbf{h}^H\mathbf{g}. \tag{10}$$

To understand the structure of parameterizations (8) and (9), one just needs to take into account that a satisfactory source extraction fulfills the following three conditions

$$\mathbf{B}\mathbf{a} = \mathbf{0} \tag{11}$$

$$\mathbf{w}^H\mathbf{Q} = \mathbf{0}^T \tag{12}$$

$$\mathbf{W}_{\mathrm{ICE}}\mathbf{A}_{\mathrm{ICE}} = \mathbf{I}_d. \tag{13}$$

The first two conditions are, in fact, included in the third one. These conditions ensure that $\mathbf{w}^H\mathbf{x} = s$ and $\mathbf{B}\mathbf{x} = \mathbf{z}$, in other words, that $\mathbf{W}_{\mathrm{ICE}}$ is de-mixing, i.e., it extracts s from $\mathbf{x}$ and separates it from $\mathbf{z}$. The ICE algebraic model can thus be written as

$$\mathbf{x} = \mathbf{A}_{\mathrm{ICE}}\mathbf{v} \tag{14}$$

where $\mathbf{v} = [s; \mathbf{z}]$.

Remark that this parameterization does not impose any restriction in the sense that the mixing matrix $\mathbf{A}$ in (1) must obey the structure given by (8) in order to extract u_1. In fact, the extraction of the background subspace is ambiguous because any transformation of that subspace does not affect the independence of the background from the SOI, and (8), resp. (9), is just a particular choice guaranteeing $\mathbf{B}\mathbf{a} = \mathbf{0}$. The ICE formulation enables us to compute the CRLB as we did in [46] for the real-valued case and Gaussian background. As compared to [46], the contribution here is that the bound is derived for the complex-valued case and it involves also the non-Gaussian background.

B. Statistical Model

The fundamental assumption of ICA/ICE states that s and $\mathbf{z}$ are independent, which means that their joint pdf can be factorized as the product of marginal pdfs. Let the pdfs of s and $\mathbf{z}$ be denoted $p_s(s)$ and $p_{\mathbf{z}}(\mathbf{z})$, respectively. Using (14), the pdf of $\mathbf{x}$ is

$$p_{\mathbf{x}}(\mathbf{x}) = p_s(\mathbf{w}^H\mathbf{x})p_{\mathbf{z}}(\mathbf{B}\mathbf{x})|\det(\mathbf{W}_{\mathrm{ICE}})|^2 \tag{15}$$

where $\det(\mathbf{W}_{\mathrm{ICE}}) = (-1)^{d-1}\gamma^{d-2}$.

C. Indeterminacies

ICE involves that same indeterminacies as ICA as the problem is solved through finding vector parameters $\mathbf{w}$ and $\mathbf{a}$ such that s and $\mathbf{z}$ are independent. It follows that any independent component of $\mathbf{x}$ could play the role of s, because of the order indeterminacy of the original components in (1). In this work, this problem can be overlooked as the CRLB analysis is local. In practice, any estimating algorithm must be properly initialized in order to extract the desired source.

The scales of s and of $\mathbf{a}$ are ambiguous in the sense that s and $\mathbf{a}$ can be substituted, respectively, by αs and $\alpha^{-1}\mathbf{a}$ with any $\alpha \neq 0$. This is know as the scaling ambiguity problem. Since Interference-to-Signal Ratio is invariant to the scaling, we can later cope with this ambiguity by fixing some scalar parameter in the mixing model. In this section, we put $\gamma = 1$. According to (15), this choice guarantees $|\det(\mathbf{W}_{\mathrm{ICE}})| = 1$, thus ensuring the non-singularity of the separating matrix.

D. Interference-to-Signal Ratio

Let $\widehat{\mathbf{w}}$ be an estimated separating vector $\mathbf{w}$. Using (7), the extracted signal is equal to $\widehat{s} = \widehat{\mathbf{w}}^H\mathbf{x} = \widehat{\mathbf{w}}^H\mathbf{a}s + \widehat{\mathbf{w}}^H\mathbf{y} = \widehat{\mathbf{w}}^H\mathbf{a}s + \widehat{\mathbf{w}}^H\mathbf{Q}\mathbf{z}$. The ISR of the signal is

$$\mathtt{ISR} = \frac{\mathrm{E}[|\widehat{\mathbf{w}}^H\mathbf{y}|^2]}{\mathrm{E}[|\widehat{\mathbf{w}}^H\mathbf{a}s|^2]} = \frac{\mathbf{q}_2^H\mathbf{C}_{\mathbf{z}}\mathbf{q}_2}{|q_1|^2\sigma_s^2} \approx \frac{1}{\sigma_s^2}\mathbf{q}_2^H\mathbf{C}_{\mathbf{z}}\mathbf{q}_2 \tag{16}$$

where $\mathbf{q}^H = [q_1, \mathbf{q}_2^H] = [\widehat{\mathbf{w}}^H\mathbf{a}, \widehat{\mathbf{w}}^H\mathbf{Q}]$, and $\mathbf{C}_{\mathbf{z}}$ stands for the covariance matrix of $\mathbf{z}$. The last approximation in (16) is valid for sufficiently small estimation error in $\widehat{\mathbf{w}}$, which is of the stochastic order of $O_p(N^{-1/2})$, having covariance of the order of $O(1/N)$. Here, $O_p(\cdot)$ represents the stochastic order symbol [47]. Note that the ISR has the same asymptotic variance, of the order of $O(1/N)$, as the corresponding CRLB. In the approximation (16), we ignore a term of the stochastic order of $O_p(N^{-3/2})$, because $\mathbf{q} \approx \mathbf{e}_1 + O_p(N^{-1/2})$, where $\mathbf{e}_1$ is the unit vector. Then, the mean ISR value reads

$$\mathrm{E}\left[\mathtt{ISR}\right] \approx \frac{1}{\sigma_s^2}\mathrm{E}\left[\mathbf{q}_2^H\mathbf{C}_{\mathbf{z}}\mathbf{q}_2\right] = \frac{1}{\sigma_s^2}\mathtt{tr}\left(\mathbf{C}_{\mathbf{z}}\mathrm{E}\left[\mathbf{q}_2\mathbf{q}_2^H\right]\right). \tag{17}$$

Hence, (17) can be written as

$$\mathrm{E}\left[\mathtt{ISR}\right] \approx \frac{1}{\sigma_s^2}\mathtt{tr}\left(\mathbf{C}_{\mathbf{z}}\mathtt{cov}\left(\mathbf{q}_2\right)\right) \tag{18}$$

where we can see that the covariance matrix of $\mathbf{q}_2$, denoted as $\mathtt{cov}\left(\mathbf{q}_2\right)$, characterizes the accuracy of $\widehat{\mathbf{w}}$. By replacing $\mathtt{cov}\left(\mathbf{q}_2\right)$ by the corresponding CRLB, we obtain the algorithm-independent Cramér-Rao-induced bound (CRIB) for ISR [4].

E. Cramér-Rao-induced Bound

Let the parameter vector be $\boldsymbol{\theta} = [\mathbf{a}; \mathbf{w}]$. In the following, we exploit a transformation rule saying that the Fisher Information Matrix (FIM) of $\boldsymbol{\theta}$, denoted as $\mathbf{F}_{\boldsymbol{\theta}}$, and the FIM of a linearly transformed version $\boldsymbol{\varphi} = \mathbf{K}\boldsymbol{\theta}$, where $\mathbf{K}$ is a non-singular matrix, are related through [48]

$$\mathbf{F}_{\boldsymbol{\varphi}} = \mathbf{K}^{-1}\mathbf{F}_{\boldsymbol{\theta}}\mathbf{K}^{-H}. \tag{19}$$

This property will be used to show that we can derive the CRIB for (18) by considering CRLB when the mixing parameters are $\mathbf{h} = \mathbf{0}$. This property is related to the *equivariance* of the BSS mixing model (1), see, e.g., [1], [49].

Now, consider the special case when $\mathbf{h} = \mathbf{g} = \mathbf{0}$, for which the parameter vector is equal to $\boldsymbol{\theta}_\mathbf{I} = [\mathbf{e}_1; \mathbf{e}_1]$. The transform between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_\mathbf{I}$ is given by

$$\boldsymbol{\theta} = \underbrace{\begin{pmatrix} \mathbf{A}_{\mathrm{ICE}} & \mathbf{0} \\ \mathbf{0} & \mathbf{W}_{\mathrm{ICE}}^H \end{pmatrix}}_{\mathbf{K}} \boldsymbol{\theta}_\mathbf{I} = \mathbf{K}\boldsymbol{\theta}_\mathbf{I} \tag{20}$$

where $\mathbf{A}_{\mathrm{ICE}}$ and $\mathbf{W}_{\mathrm{ICE}}$ are, respectively, given by (8) and (9). According to (19), it holds that

$$\mathbf{F}_{\boldsymbol{\theta}} = \mathbf{K}\mathbf{F}_{\boldsymbol{\theta}_\mathbf{I}}\mathbf{K}^H. \tag{21}$$

Similarly, we can consider a transformed parameter vector

$$\boldsymbol{\theta}_\mathbf{q} = \underbrace{\begin{pmatrix} \mathbf{W}_{\mathrm{ICE}} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_{\mathrm{ICE}}^H \end{pmatrix}}_{\mathbf{K}^{-1}} \begin{pmatrix} \mathbf{a} \\ \mathbf{w} \end{pmatrix} = \mathbf{K}^{-1}\boldsymbol{\theta} \tag{22}$$

and it holds that $\mathbf{F}_{\boldsymbol{\theta}_\mathbf{q}} = \mathbf{K}^{-1}\mathbf{F}_{\boldsymbol{\theta}}\mathbf{K}^{-H}$, which, together with (21), results in

$$\mathbf{F}_{\boldsymbol{\theta}_\mathbf{q}} = \mathbf{F}_{\boldsymbol{\theta}_\mathbf{I}}. \tag{23}$$

Hence, from (23) it follows that the CRIB for (18) can be obtained by replacing $\mathtt{cov}\left(\mathbf{q}_2\right)$ by the corresponding CRLB, which is equal to the CRLB for the unbiased estimation of $\mathbf{h}$ when its true value is $\mathbf{h} = \mathbf{0}$. Finally,

$$\mathrm{E}\left[\mathtt{ISR}\right] \approx \frac{1}{\sigma_s^2}\mathtt{tr}(\mathbf{C}_\mathbf{z}\mathtt{cov}(\widehat{\mathbf{h}})) \geq \frac{1}{\sigma_s^2}\mathtt{tr}\left(\mathbf{C}_\mathbf{z}\mathtt{CRLB}\left(\mathbf{h}\right)|_{\mathbf{h}=\mathbf{0}}\right) \tag{24}$$

where $\mathtt{CRLB}(\mathbf{h})|_{\mathbf{h}=\mathbf{0}}$ denotes the diagonal block of the inverse matrix of the FIM corresponding to the parameter vector $\mathbf{h}$ when $\mathbf{h} = \mathbf{0}$. The inequality between the mean ISR and the corresponding lower bound is approximate, but its leading term is the same on both sides. Ignoring higher-order terms is common in the literature.

F. Fisher Information Matrix

To compute the CRLB, we use the approach for the complex-valued parameters described in [48]. By putting $\gamma = 1$, as justified in Section III-C, the only free parameters of the mixing model (14) are $\mathbf{h}$ and $\mathbf{g}$, so let the parameter vector be $\boldsymbol{\theta} = [\mathbf{h}; \mathbf{g}]$. According to [48], for any unbiased estimator of $\boldsymbol{\theta}$, it holds that

$$\mathtt{cov}\left(\boldsymbol{\theta}\right) \succeq \mathcal{J}^{-1}\left(\boldsymbol{\theta}\right) = \mathtt{CRLB}\left(\boldsymbol{\theta}\right), \tag{25}$$

where $\mathcal{J}\left(\boldsymbol{\theta}\right)$ is the FIM, and $\mathbf{C} \succeq \mathbf{D}$ means that $\mathbf{C} - \mathbf{D}$ is a positive semi-definite matrix. $\mathcal{J}\left(\boldsymbol{\theta}\right)$ can be partitioned as

$$\mathcal{J}\left(\boldsymbol{\theta}\right) = \begin{pmatrix} \mathbf{F} & \mathbf{P} \\ \mathbf{P}^* & \mathbf{F}^* \end{pmatrix}, \tag{26}$$

where

$$\mathbf{F} = \mathrm{E}\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^*}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^*}\right)^H\right], \quad \mathbf{P} = \mathrm{E}\left[\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^*}\left(\frac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}^*}\right)^T\right] \tag{27}$$

and where the derivatives in (27) are defined according to Wirtinger calculus. $\mathcal{L}(\cdot)$ denotes the log-likelihood function of (15), namely,

$$\mathcal{L}(\mathbf{h}, \mathbf{g}|\mathbf{x}) = \log p_s(\mathbf{w}^H\mathbf{x}) + \log p_\mathbf{z}(\mathbf{B}\mathbf{x}). \tag{28}$$

The derivatives of the log-likelihood function (28) are as follows:

$$\frac{\partial \mathcal{L}(\mathbf{x}|\boldsymbol{\theta})}{\partial \mathbf{g}^*}\Big|_{\mathbf{h}=0} = -\boldsymbol{\psi}_\mathbf{z}(\mathbf{z})s^* \tag{29}$$

$$\frac{\partial \mathcal{L}(\mathbf{x}|\boldsymbol{\theta})}{\partial \mathbf{h}^*}\Big|_{\mathbf{h}=0} = \psi_s^*(s)\mathbf{z} \tag{30}$$

where $\psi_s(s) = -\frac{\partial \ln p_s(s,s^*)}{\partial s^*}$ and $\boldsymbol{\psi}_\mathbf{z}(\mathbf{z}) = -\frac{\partial \ln p_\mathbf{z}(\mathbf{z},\mathbf{z}^*)}{\partial \mathbf{z}^*}$ are the score functions. Using (29),(30), $\mathbf{F}$ in (27) is calculated as

$$\mathbf{F} = \begin{pmatrix} \sigma_s^2\boldsymbol{\kappa}_\mathbf{z} & -\mathbf{I}_{d-1} \\ -\mathbf{I}_{d-1} & \kappa_s\mathbf{C}_\mathbf{z} \end{pmatrix}, \tag{31}$$

where

$$\kappa_s = \mathrm{E}[|\psi(s)|^2], \tag{32}$$

$$\sigma_s^2 = \mathrm{E}[|s|^2], \tag{33}$$

$$\boldsymbol{\kappa}_\mathbf{z} = \mathrm{E}\left[\boldsymbol{\psi}_\mathbf{z}(\mathbf{z})\boldsymbol{\psi}_\mathbf{z}^H(\mathbf{z})\right]. \tag{34}$$

Now, we describe the computation of $\mathbf{P}$ in (26). Let $\mathbf{P}$ be partitioned as

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}_{\mathbf{g},\mathbf{g}} & \mathbf{P}_{\mathbf{g},\mathbf{h}} \\ \mathbf{P}_{\mathbf{g},\mathbf{h}}^T & \mathbf{P}_{\mathbf{h},\mathbf{h}} \end{pmatrix}. \tag{35}$$

Then,

$$\mathbf{P}_{\mathbf{g},\mathbf{g}} = \mathrm{E}\left[\boldsymbol{\psi}_\mathbf{z}(\mathbf{z})\boldsymbol{\psi}_\mathbf{z}^T(\mathbf{z})\right]\mathrm{E}\left[s^{*2}\right] \tag{36}$$

$$\mathbf{P}_{\mathbf{h},\mathbf{h}} = \mathrm{E}\left[\psi_s^*(s)^2\right]\mathrm{E}\left[\mathbf{z}\mathbf{z}^T\right] \tag{37}$$

$$\mathbf{P}_{\mathbf{g},\mathbf{h}} = \mathbf{0}. \tag{38}$$

G. Circular Sources

In general, the analytic computation of the inverse matrix of (26) is not tractable. Therefore, we investigate two special cases in the following subsections.

Here, we assume that s and $\mathbf{z}$ have general *circular* pdf. Under this assumption, the FIM (26) obtains the block-diagonal structure, because $\mathbf{P}_{\mathbf{h},\mathbf{h}} = \mathbf{P}_{\mathbf{g},\mathbf{h}} = \mathbf{0}$ due to the circularity of $\mathbf{z}$ and $\mathbf{P}_{\mathbf{g},\mathbf{g}} = \mathbf{0}$ due to the circularity of s, and, then,

$$\mathcal{J}\left(\boldsymbol{\theta}\right) = \begin{pmatrix} \sigma_s^2\boldsymbol{\kappa}_\mathbf{z}^{-1} & -\mathbf{I}_{d-1} & \mathbf{O} \\ -\mathbf{I}_{d-1} & \kappa_s\mathbf{C}_\mathbf{z} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{F}^* \end{pmatrix}. \tag{39}$$

$\mathtt{CRLB}(\mathbf{h})|_{\mathbf{h}=\mathbf{0}}$ is obtained as the upper right diagonal block of the inverse matrix of (39), which reads

$$\mathtt{CRLB}(\mathbf{h})|_{\mathbf{h}=\mathbf{0}} = \left(\kappa_s\mathbf{C}_\mathbf{z} - \frac{1}{\sigma_s^2}\boldsymbol{\kappa}_\mathbf{z}^{-1}\right)^{-1}. \tag{40}$$

Applying the transformation theorem in (34), it can be shown that, for $\tilde{\mathbf{z}} = \mathbf{T}\mathbf{z}$, it holds that

$$\boldsymbol{\kappa}_\mathbf{z} = \mathbf{T}\boldsymbol{\kappa}_{\tilde{\mathbf{z}}}\mathbf{T}^H \tag{41}$$

where $\mathbf{T}$ is a non-singular transformation matrix. By taking $\mathbf{T} = \mathbf{C}_\mathbf{z}^{-\frac{1}{2}}$, which is a matrix satisfying that $\mathbf{C}_\mathbf{z}^{-\frac{1}{2}}\mathbf{C}_\mathbf{z}^{-\frac{1}{2}} =$

$\mathbf{C}_{\mathbf{z}}^{-1}$, then $\boldsymbol{\kappa}_{\tilde{\mathbf{z}}}$ corresponds to the statistic of uncorrelated and unit-scaled $\mathbf{z}$. Hence, (40) can be written as

$$\mathtt{CRLB}(\mathbf{h})|_{\mathbf{h}=\mathbf{0}} = \mathbf{C}_{\mathbf{z}}^{-\frac{1}{2}} \left(\kappa_s \mathbf{I}_{d-1} - \frac{1}{\sigma_s^2} \boldsymbol{\kappa}_{\tilde{\mathbf{z}}}^{-1} \right)^{-1} \mathbf{C}_{\mathbf{z}}^{-\frac{1}{2}}. \quad (42)$$

By putting (42) into (24), the CRIB for ISR, when considering N observations, is

$$\mathrm{E}\left[\mathtt{ISR}\right] \geq \frac{1}{N} \frac{1}{\sigma_s^2} \mathtt{tr} \left[\left(\kappa_s \mathbf{I}_{d-1} - \frac{1}{\sigma_s^2} \boldsymbol{\kappa}_{\tilde{\mathbf{z}}}^{-1} \right)^{-1} \right]. \quad (43)$$

Next, we can use the identity (41) again by considering $\mathbf{T}$ such that elements of $\mathbf{T}\tilde{\mathbf{z}}$ are statistically independent random variables. Since elements of $\tilde{\mathbf{z}}$ are uncorrelated and normalized, such $\mathbf{T}$ must be unitary, i.e., $\mathbf{T}\mathbf{T}^H = \mathbf{I}_{d-1}$. Also, provided that all but one components in the original model (1) are non-Gaussian, the entire mixture is separable, so $\mathbf{T}\tilde{\mathbf{z}}$ must be equal to $\mathbf{u}_2$ up to their order and scales. Without any loss of generality, we can assume that $\mathbf{T}$ is such that $\mathbf{T}\tilde{\mathbf{z}} = \mathbf{u}_2$ and that $\mathbf{u}_2$ have unit variance. Then, $\boldsymbol{\kappa}_{\mathbf{T}\tilde{\mathbf{z}}}$ is diagonal having diagonal elements equal to $\overline{\kappa}_2, \ldots, \overline{\kappa}_d$, and (43) simplifies to

$$\mathrm{E}\left[\mathtt{ISR}\right] \geq \frac{1}{N} \sum_{j=2}^{d} \frac{\overline{\kappa}_j}{\sigma_s^2 \kappa_s \overline{\kappa}_j - 1}. \quad (44)$$

This bound corresponds with (5) for $i = 1$ since $\sigma_s^2 \kappa_s = \overline{\kappa}_s = \overline{\kappa}_1$, which means that the same extraction accuracy can be achieved by ICE as by ICA. It should be, however, noted that the multivariate score function $\psi_{\mathbf{z}}(\cdot)$ must be known for maximum likelihood estimation to be carried out [50].

In our considerations, we can go also slightly beyond the standard ICA. Let the observed signals obey the model (14) but not (1), that is, there need not exist $\mathbf{T}$ such that $\mathbf{T}\tilde{\mathbf{z}}$ are independent (no independent components $u_2, \ldots, u_d$ are assumed, only the independence between s and $\mathbf{z}$). Since $\boldsymbol{\kappa}_{\tilde{\mathbf{z}}}$ is positive definite, we can consider its decomposition

$$\boldsymbol{\kappa}_{\tilde{\mathbf{z}}} = \mathbf{U}\mathbf{D}\mathbf{U}^H \quad (45)$$

where $\mathbf{U}^H$ is the unitary matrix of eigenvectors of $\boldsymbol{\kappa}_{\tilde{\mathbf{z}}}$, and $\mathbf{D}$ is diagonal with diagonal entries denoted as $\omega_2, \ldots, \omega_d$. Then, the CRIB presents a form similar to (44):

$$\mathrm{E}\left[\mathtt{ISR}\right] \geq \frac{1}{N} \sum_{j=2}^{d} \frac{\omega_j}{\sigma_s^2 \kappa_s \omega_j - 1}. \quad (46)$$

H. Circular Gaussian Background

Here, we assume that s can be arbitrary non-circular and non-Gaussian while $\mathbf{z}$ is circular Gaussian. Under this assumption, $\mathbf{P}_{\mathbf{h},\mathbf{h}} = \mathbf{0}$, and since $\boldsymbol{\kappa}_{\mathbf{z}} = \mathbf{C}_{\mathbf{z}}^{-1}$, also $\mathbf{P}_{\mathbf{g},\mathbf{g}} = \mathbf{0}$ thanks to the circularity of $\mathbf{z}$. The FIM thus shows a structure similar to (39), namely,

$$\mathcal{J}(\boldsymbol{\theta}) = \begin{pmatrix} \sigma_s^2 \mathbf{C}_{\mathbf{z}}^{-1} & -\mathbf{I}_{d-1} & \mathbf{O} \\ -\mathbf{I}_{d-1} & \kappa_s \mathbf{C}_{\mathbf{z}} & \mathbf{O} \\ \mathbf{O} & \mathbf{O} & \mathbf{F}^* \end{pmatrix}. \quad (47)$$

Hence,

$$\mathtt{CRLB}(\mathbf{h})|_{\mathbf{h}=\mathbf{0}} = \left(\kappa_s \mathbf{C}_{\mathbf{z}} - \frac{1}{\sigma_s^2} \mathbf{C}_{\mathbf{z}} \right)^{-1} = \frac{\sigma_s^2}{\kappa_s \sigma_s^2 - 1} \mathbf{C}_{\mathbf{z}}^{-1}. \quad (48)$$

Therefore, for N observations, the CRIB for ISR says that

$$\mathrm{E}\left[\mathtt{ISR}\right] \geq \frac{1}{N} \frac{d-1}{\overline{\kappa}_s - 1}. \quad (49)$$

This result confirms the asymptotic bound given by (6) for complex-valued non-circular SOI.

IV. CRLB for Piecewise Determined Mixing

We now turn to the piecewise determined mixtures, in general, described by (4). To deal with this model, we begin by proposing a generalization of the ICE concept as follows.

A. Algebraic Model

Without any further assumption, (4) corresponds to a sequential application of the standard mixing model, which is straightforward for on-line signal processing but does not bring any advantage. Therefore, we propose special parameterizations useful for the BSE problem assuming that the SOI is active in all blocks and some mixing parameters related to the SOI are joint to all the blocks. Specifically, we parametrize $\mathbf{A}^1, \ldots, \mathbf{A}^M$ similarly to (8) and consider two special variants:

$$\mathbf{A}^m_{\mathrm{CMV}} = \begin{pmatrix} \gamma & (\mathbf{h}^m)^H \\ \mathbf{g} & \frac{1}{\gamma}\left(\mathbf{g}(\mathbf{h}^m)^H - \mathbf{I}_{d-1}\right) \end{pmatrix} \quad (50)$$

$$\mathbf{A}^m_{\mathrm{CSV}} = \begin{pmatrix} \gamma^m & \mathbf{h}^H \\ \mathbf{g}^m & \frac{1}{\gamma}\left(\mathbf{g}^m \mathbf{h}^H - \mathbf{I}_{d-1}\right). \end{pmatrix}. \quad (51)$$

The models will be referred to as constant mixing vector (CMV) and constant separating vector (CSV), respectively, because in CMV the mixing vectors $\mathbf{a}^1, \ldots, \mathbf{a}^M$ are constant over blocks and are equal to $\mathbf{a}$ while in CSV the separating vectors $\mathbf{w}^1, \ldots, \mathbf{w}^M$ are all equal to $\mathbf{w}$. CMV is useful for situations where the SOI is a static source while the background is varying. CSV involves a moving SOI (varying mixing vector) under the assumption that a constant separating vector such that extracts the signal from all blocks exists. These models have been considered for the first time in [12], where they were applied to blind audio source extraction. This paper provides their theoretical analysis through the CRLB theory.

B. Interference-to-Signal Ratio

For simplicity, let the number of available samples in each of M blocks be the same, equal to N_b. It holds that $M \cdot N_b = N$. The variance of the SOI and the covariance matrix of the background signals in the mth block will be denoted by $\sigma_{s^m}^2$ and $\mathbf{C}_{\mathbf{z}^m}$, respectively.

Let $\widehat{\mathbf{w}}^m$ be an estimated separating vector for the mth block, $m = 1, \ldots, M$. The ISR of the extracted signal evaluated over the entire data is equal to

$$\mathtt{ISR} = \frac{\sum_{m=1}^{M} \mathrm{E}[|(\widehat{\mathbf{w}}^m)^H \mathbf{y}^m|^2]}{\sum_{m=1}^{M} \mathrm{E}[|(\widehat{\mathbf{w}}^m)^H \mathbf{a}^m s^m|^2]} =$$
$$= \frac{\sum_{m=1}^{M} (\mathbf{q}_2^m)^H \mathbf{C}_{\mathbf{z}^m} \mathbf{q}_2^m}{\sum_{m=1}^{M} |q_1^m|^2 \sigma_{s^m}^2} = \frac{\sum_{m=1}^{M} \mathtt{tr}\left(\mathbf{C}_{\mathbf{z}^m} \mathbf{q}_2^m (\mathbf{q}_2^m)^H \right)}{\sum_{m=1}^{M} |q_1^m|^2 \sigma_{s^m}^2}, \quad (52)$$

where $(\mathbf{q}^m)^H = [q_1^m, (\mathbf{q}_2^m)^H] = [(\widehat{\mathbf{w}}^m)^H\mathbf{a}^m, (\widehat{\mathbf{w}}^m)^H\mathbf{Q}^m]$. Assuming "small" estimation errors, i.e., $\mathbf{q}^m \approx \mathbf{e}_1$, similar approximation to that in (16) gives

$$\mathtt{ISR} \approx \frac{1}{\sum_{m=1}^M \sigma_{s^m}^2} \sum_{m=1}^M \mathtt{tr}\Big(\mathbf{C}_{\mathbf{z}^m}\mathbf{q}_2^m(\mathbf{q}_2^m)^H\Big). \tag{53}$$

Using the equivariance property described in Section III-E, the CRIB is, in general, obtained through

$$\mathrm{E}\left[\mathtt{ISR}\right] \geq \frac{1}{\sum_{m=1}^M \sigma_{s^m}^2} \mathtt{tr}\left(\sum_{m=1}^M \mathbf{C}_{\mathbf{z}^m}\mathtt{CRLB}\left(\mathbf{h}^m\right)|_{\substack{\mathbf{h}^m=0\\ \mathbf{g}^m=0}}\right). \tag{54}$$

C. Blockwise ICE

To extract the SOI from each block of data (4), the ICE approach can be used. Then, the mixing and separating vectors are estimated as parameters that are independent of the other blocks. We will refer to this approach as block ICE (BICE).

Assuming that the background is circular Gaussian, the CRIB for BICE follows from the results of Section III-H. By putting (48) into (54) and using the fact all data are independently distributed, the CRIB is given by

$$\mathrm{E}[\mathtt{ISR}] \geq \frac{1}{N_b}\frac{d-1}{\sum_{m=1}^M \sigma_{s^m}^2} \sum_{m=1}^M \frac{\sigma_{s^m}^2}{\kappa_{s_m}\sigma_{s^m}^2 - 1}. \tag{55}$$

It is worth comparing this bound with CRIBs derived for the CMV and CSV models given by (50) and (51), respectively, which is the subject of the following subsections.

D. Constant Mixing Vector

In the CMV model, $\mathbf{a}$ is constant over M blocks while the separating vector can be varying from block to block. Therefore, there are $M(d-1)+d$ free parameters. The scaling ambiguity can be resolved by putting $\gamma = 1$, which is the first element of $\mathbf{a}$, so there are finally $(M+1)(d-1)$ free (complex-valued) parameters in the mixing model represented by parameter vectors $\mathbf{g}$ and $\mathbf{h} = [\mathbf{h}^1; \ldots; \mathbf{h}^M]$.

From (28), it follows that the log-likelihood function for one sample data of the mth block is given by

$$\mathcal{L}^m(\mathbf{x}^m|\mathbf{g},\mathbf{h}) = \log p_{s^m}\left((\mathbf{w}^m)^H\mathbf{x}^m\right) + \log p_{\mathbf{z}^m}(\mathbf{B}\mathbf{x}^m). \tag{56}$$

Since the data are i.i.d. inside each block and independently distributed among the blocks, the log-likelihood function of the entire batch of data is equal to $N_b\sum_{m=1}^M \mathcal{L}^m(\mathbf{x}^m|\mathbf{g},\mathbf{h})$.

The derivatives of (56) are computed similarly to (29) and (30), that is,

$$\boldsymbol{\nabla}_{\mathbf{g}}^m \overset{\text{def.}}{=} \left.\frac{\partial\mathcal{L}^m(\mathbf{x}^m|\mathbf{g},\mathbf{h})}{\partial\mathbf{g}^*}\right|_{\mathbf{h}=\mathbf{0}} = -\boldsymbol{\psi}_{\mathbf{z}^m}s^{m*}, \tag{57}$$

$$\boldsymbol{\nabla}_{\mathbf{h}}^{m,n} \overset{\text{def.}}{=} \left.\frac{\partial\mathcal{L}^m(\mathbf{x}^m|\mathbf{g},\mathbf{h})}{\partial\mathbf{h}^{n*}}\right|_{\mathbf{h}=\mathbf{0}} = \delta_{n,m}\psi_{s^m}^*\mathbf{z}^m, \tag{58}$$

where $\boldsymbol{\psi}_{\mathbf{z}^m} = -\frac{\partial \ln p_{\mathbf{z}^m}}{\partial \mathbf{z}^*}$, $\psi_{s^m} = -\frac{\partial \ln p_{s^m}}{\partial s^*}$, and $\delta_{n,m}$ stands for the Kronecker delta.

Now, the FIM of data from all blocks is a square matrix of dimension $2(m+1)(d-1)$ consisting of $(m+1)\times(m+1)$ blocks each of dimension $(d-1)\times(d-1)$. Let $\boldsymbol{\nabla}^m = [\boldsymbol{\nabla}_{\mathbf{g}}^m; \boldsymbol{\nabla}_{\mathbf{h}}^{m,1}; \ldots; \boldsymbol{\nabla}_{\mathbf{h}}^{m,M}]$. The FIM has the structure

$$\mathcal{J}(\mathbf{g},\mathbf{h}) = N_b\sum_{m=1}^M \mathcal{J}^m(\mathbf{g},\mathbf{h}) = N_b\begin{pmatrix}\mathbf{F} & \mathbf{P}\\ \mathbf{P}^* & \mathbf{F}^*\end{pmatrix}, \tag{59}$$

where

$$\mathcal{J}^m(\mathbf{g},\mathbf{h}) = \begin{pmatrix}\mathbf{F}^m & \mathbf{P}^m\\ \mathbf{P}^{m*} & \mathbf{F}^{m*}\end{pmatrix} \tag{60}$$

is the FIM for one sample of the mth block, and

$$\mathbf{F}^m = \mathrm{E}\left[\boldsymbol{\nabla}^m(\boldsymbol{\nabla}^m)^H\right], \qquad \mathbf{P}^m = \mathrm{E}\left[\boldsymbol{\nabla}^m(\boldsymbol{\nabla}^m)^T\right]. \tag{61}$$

Then the blocks of (59) are, respectively, equal to

$$\mathbf{F} = \begin{pmatrix}\sum_{m=1}^M \boldsymbol{\kappa}_{\mathbf{z}^m}\sigma_{s^m}^2 & -\mathbf{I}_{d-1} & \ldots & -\mathbf{I}_{d-1}\\ -\mathbf{I}_{d-1} & \kappa_s^1\mathbf{C}_{\mathbf{z}^1} & & \mathbf{0}\\ \vdots & \mathbf{0} & \ddots & \\ -\mathbf{I}_{d-1} & & & \kappa_s^M\mathbf{C}_{\mathbf{z}^M}\end{pmatrix}, \tag{62}$$

and $\mathbf{P}$ is the diagonal matrix

$$\mathbf{P} = \mathrm{diag}\Big(\sum_{m=1}^M \mathrm{E}[\boldsymbol{\psi}_{\mathbf{z}^m}^2]\mathrm{E}[(s^{m*})^2], \mathrm{E}[(\boldsymbol{\psi}_{s^1}^*)^2]\mathrm{E}[(\mathbf{z}^1)^2], \ldots, \mathrm{E}[(\boldsymbol{\psi}_{s^M}^*)^2]\mathrm{E}[(\mathbf{z}^M)^2]\Big) \tag{63}$$

where $\kappa_{s^m} = \mathrm{E}[|\boldsymbol{\psi}_{s^m}|^2]$, $\boldsymbol{\kappa}_{\mathbf{z}^m} = \mathrm{E}[\boldsymbol{\psi}_{\mathbf{z}^m}\boldsymbol{\psi}_{\mathbf{z}^m}^H]$, $\sigma_{s^m}^2 = \mathrm{E}[|s^m|^2]$, $\mathbf{C}_{\mathbf{z}^m} = \mathrm{E}[\mathbf{z}^m(\mathbf{z}^m)^H]$.

For the sake of simplicity, we will consider only the special case when the background is circular Gaussian. Then, similar simplifications to those in Section III-H hold, $\mathbf{P} = \mathbf{0}$, $\boldsymbol{\kappa}_{\mathbf{z}^m} = \mathbf{C}_{\mathbf{z}^m}^{-1}$, and the block of $\mathcal{J}^{-1}$ corresponding to $\mathbf{h}^m$ is

$$\mathtt{CRLB}(\mathbf{h}^m)|_{\mathbf{h}=0} = \frac{1}{N_b}\left\{\frac{1}{\kappa_{s^m}}\mathbf{C}_{\mathbf{z}^m}^{-1} + \frac{1}{\kappa_{s^m}}\mathbf{C}_{\mathbf{z}^m}^{-1}\left(\sum_{i=1}^M \frac{\sigma_{s^i}^2\kappa_{s^i}-1}{\kappa_{s^i}}\mathbf{C}_{\mathbf{z}^i}^{-1}\right)^{-1}\frac{1}{\kappa_{s^m}}\mathbf{C}_{\mathbf{z}^m}^{-1}\right\}. \tag{64}$$

By combining (54) and (64), the CRIB says that

$$\mathrm{E}\left[\mathtt{ISR}\right] \geq \frac{1}{N_b\sum_{m=1}^M \sigma_{s^m}^2}\sum_{m=1}^M \frac{1}{\kappa_{s^m}} \times \mathtt{tr}\left(\mathbf{I}_{d-1} + \left(\sum_{i=1}^M \frac{\overline{\kappa}_{s^i}-1}{\kappa_{s^i}}\mathbf{C}_{\mathbf{z}^i}^{-1}\right)^{-1}\frac{1}{\kappa_{s^m}}\mathbf{C}_{\mathbf{z}^m}^{-1}\right). \tag{65}$$

E. Constant Separating Vector

In the CSV mixing model (51), $\mathbf{w}$ is constant over the blocks while the mixing vector can be varying. Therefore, the scaling ambiguity can be resolved by putting $\beta = 1$ while considering $\gamma^1, \ldots, \gamma^M$ as dependent variables, where by (10) it follows that $\gamma^m = 1 - \mathbf{h}^H\mathbf{g}^m$. The free parameter vectors of the model are $\mathbf{g} = [\mathbf{g}^1; \ldots; \mathbf{g}^M]$ and $\mathbf{h}$.

Using (15), the log-likelihood function for one sample of the mth block is

$$\mathcal{L}^m(\mathbf{x}^m|\mathbf{g},\mathbf{h}) = \log p_{s^m}\left(\mathbf{w}^H\mathbf{x}^m\right) + \log p_{\mathbf{z}}^m(\mathbf{B}^m\mathbf{x}^m) + + 2(d-2)\log\left|1-\mathbf{h}^H\mathbf{g}^m\right|, \tag{66}$$

where we use the identity $\det(\mathbf{W}_{\mathrm{ICE}}) = (-1)^{d-1}(1 - \mathbf{h}^H\mathbf{g}^m)^{d-2}$.

The structure of the FIM is the same as for the CMV model, described by (59)-(61). The blocks of (59) are given by

$$\mathbf{F} = \begin{pmatrix} \boldsymbol{\kappa}_{\mathbf{z}^1}\sigma_{s^1}^2 & & \mathbf{0} & -\mathbf{I}_{d-1} \\ & \ddots & & \vdots \\ \mathbf{0} & & \boldsymbol{\kappa}_{\mathbf{z}^M}\sigma_{s^M}^2 & -\mathbf{I}_{d-1} \\ -\mathbf{I}_{d-1} & \dots & -\mathbf{I}_{d-1} & \sum_{m=1}^{M}\kappa_s^m\mathbf{C}_{\mathbf{z}^m} \end{pmatrix}, \quad (67)$$

and $\mathbf{P}$ is diagonal

$$\mathbf{P} = \mathrm{diag}\big(\mathrm{E}[(\boldsymbol{\psi}_{\mathbf{z}^1})^2]\mathrm{E}[(s^{1^*})^2], \dots, \mathrm{E}[(\boldsymbol{\psi}_{\mathbf{z}^M})^2]\mathrm{E}[(s^{M^*})^2], \sum_{m=1}^{M}\mathrm{E}[(\boldsymbol{\psi}_{s^m}^*)^2]\mathrm{E}[(\mathbf{z}^m)^2]\big). \quad (68)$$

Here, we also consider only the special case that the background is circular Gaussian, for which $\mathbf{P} = \mathbf{0}$, $\boldsymbol{\kappa}_{\mathbf{z}^m} = \mathbf{C}_{\mathbf{z}^m}^{-1}$. Then, $\mathtt{CRLB}(\mathbf{h})|_{\mathbf{h}=\mathbf{g}=0}$ is obtained as the block of the inverse matrix of FIM corresponding to the lower right-corner block of $\mathbf{F}$, which gives

$$\mathtt{CRLB}(\mathbf{h})|_{\mathbf{h}=\mathbf{g}=0} = \frac{1}{N_b}\left(\sum_{m=1}^{M}\kappa_s^m\mathbf{C}_{\mathbf{z}^m} - \frac{1}{\sigma_{s^m}^2}\mathbf{C}_{\mathbf{z}^m}\right)^{-1}. \quad (69)$$

By putting this result in (54), the CRIB says that

$$\mathrm{E}\left[\mathtt{ISR}\right] \geq \frac{1}{N_b\sum_{m=1}^{M}\sigma_{s^m}^2}\times \mathtt{tr}\left(\left(\sum_{m=1}^{M}\frac{\overline{\kappa}_{s^m}-1}{\sigma_{s^m}^2}\mathbf{C}_{\mathbf{z}^m}\right)^{-1}\sum_{m=1}^{M}\mathbf{C}_{\mathbf{z}^m}\right). \quad (70)$$

F.

In this subsection, we show that the CRIB expressions for the ISR in the CSV and CMV models (65) and (70) remain valid also in the simpler scenario composed of real-valued signals and mixing parameters. In this case, the pdf of mixed signals in (15) becomes

$$p_{\mathbf{x}}(\mathbf{x}) = p_s(\mathbf{w}^T\mathbf{x})p_{\mathbf{z}}(\mathbf{B}\mathbf{x})|\det(\mathbf{W}_{\mathrm{ICE}})|. \quad (71)$$

Next, the fact that the Gaussian pdf of background is real-valued has to be taken into account. Then, the FIM in (59) reduces to

$$\mathcal{J}(\mathbf{g},\mathbf{h}) = N_b\mathbf{F}. \quad (72)$$

Matrix $\mathbf{F}$ can be shown to have exactly the same form as (62) for the CMV model and as (67) for the CSV model. Therefore, formally the same CRLBs are finally obtained in the real-valued scenario.

V. Discussion

The expressions in brackets in (65) and (70) subject to the matrix inverse operation are non-negative combinations of positive definite matrices ($\mathbf{C}_{\mathbf{z}_m}^{-1}$ or $\mathbf{C}_{\mathbf{z}_m}$). It follows that the sums are also positive definite unless all coefficients of the linear combinations are zero. The latter case appears only if $\overline{\kappa}_{s^m} = 1$ for all m, that is, when the SOI is Gaussian on all blocks. Otherwise, the obtained CRIBs are all finite.

In the following, we discuss several special cases in order to compare the derived bounds.

1) Only one block: When $M = 1$, the piecewise determined models coincide with the standard ICE model. The reader can easily verify that, for this particular case, the bounds given by (49), (55), (65) and (70) coincide as well.

In further discussions, we will assume the piecewise model with $M > 1$.

2) An i.i.d. SOI: When the SOI has the same pdf (and also variance) in all blocks, we can denote $\kappa_s^m = \kappa_s$ and $\sigma_{s^m}^2 = \sigma_s^2$ since these statistics become independent of m. Then, the CRIBs (55), (65) and (70) can be, respectively, written in the form

$$\text{BICE:}\quad \mathrm{E}[\mathtt{ISR}] \geq \frac{M}{N}\frac{d-1}{\overline{\kappa}_s - 1} \quad (73)$$

$$\text{CMV:}\quad \mathrm{E}[\mathtt{ISR}] \geq \frac{d-1}{N}\left(\frac{1}{\overline{\kappa}_s - 1} + \frac{M-1}{\overline{\kappa}_s}\right) \quad (74)$$

$$\text{CSV:}\quad \mathrm{E}[\mathtt{ISR}] \geq \frac{1}{N}\frac{d-1}{\overline{\kappa}_s - 1} \quad (75)$$

A necessary condition for the identifiability of these models is that $\overline{\kappa}_s > 1$, which means that the SOI must have non-Gaussian pdf. The CRIB for BICE is always higher than those for CSV and CMV, which is caused by the higher complexity of BICE. CSV and CMV take advantage of the joint parameters.

3) SOI with varying variance: Let the variance of the SOI be changing from block to block while the normalized pdf of the SOI be constant. It means that $\sigma_{s^m}^2$ depends on m while $\kappa_{s^m}\sigma_{s^m}^2 = \overline{\kappa}_s$ is constant over the blocks. Then, the CRIBs can be written as

$$\text{BICE:}\quad \mathrm{E}[\mathtt{ISR}] \geq \frac{M}{N}\frac{d-1}{\overline{\kappa}_s - 1}, \quad (76)$$

$$\text{CMV:}\quad \mathrm{E}[\mathtt{ISR}] \geq \frac{M(d-1)}{N\overline{\kappa}_s} + \frac{M}{N\overline{\kappa}_s(\overline{\kappa}_s - 1)}T_{\mathrm{CMV}}, \quad (77)$$

$$\text{CSV:}\quad \mathrm{E}[\mathtt{ISR}] \geq \frac{M}{N(\overline{\kappa}_s - 1)}T_{\mathrm{CSV}}, \quad (78)$$

where

$$T_{\mathrm{CMV}} = \mathtt{tr}\left(\sum_{m=1}^{M}\frac{\sigma_{s^m}^2}{\sum_{j=1}^{M}\sigma_{s^j}^2}\left(\sum_{i=1}^{M}\mathbf{S}_i\right)^{-1}\mathbf{S}_m\right), \quad (79)$$

$$T_{\mathrm{CSV}} = \mathtt{tr}\left(\frac{1}{\sum_{j=1}^{M}\sigma_{s^j}^2}\left(\sum_{i=1}^{M}\frac{1}{\sigma_{s^i}^2}\mathbf{C}_{\mathbf{z}^m}\right)^{-1}\sum_{m=1}^{M}\mathbf{C}_{\mathbf{z}^m}\right), \quad (80)$$

where $\mathbf{S}_m = \sigma_{s^m}^2\mathbf{C}_{\mathbf{z}^m}^{-1}$.

The bound given by (76) coincides with (73), which means that the dynamic envelop of the SOI does not have any influence on the achievable performance when ICE is independently applied to each block. By comparing (77) with (74) and (78) with (75), we obtain more interesting results.

It can be easily shown that

$$T_{\mathrm{CMV}} \leq d-1, \tag{81}$$

$$T_{\mathrm{CSV}} \leq \frac{d-1}{M}. \tag{82}$$

It follows that the bound (78) is always lower than the one given by (75), moreover, the equality of the bounds holds if and only if $\sigma^2_{s^m}$ is constant. It means that the non-stationarity of the SOI improves the blind extraction under the CSV model. This is not that surprising because similar conclusions follow from Cramér-Rao analyses for the standard BSS models that involve signals' non-stationarity. There, more dynamical signals improve the achievable separation accuracy; see, e.g., [27], [30], [51].

For the upper limit in (81), the bound coincides with (76). It means that the achievable ISR by CMV is never worse than that by BICE. Next, it is easily seen that $T_{\mathrm{CMV}} = \frac{d-1}{M}$ when $\sigma^2_{s^m}$ is constant, for which case the bound obviously coincides with (74). It means that the nonstationarity of the SOI can improve as well as worsen the extraction accuracy under the CMV model! A deeper analysis can show that the latter case is more typical, because, for the improved accuracy, background must be vanishing or badly conditioned in some block. When $\mathbf{C}_{\mathbf{z}^m}$ is constant over blocks, then $T_{\mathrm{CMV}} = \frac{d-1}{M}$ is the lower bound, and the SOI's nonstationarity can only worsen the achievable accuracy under the CMV model.

4) All but one blocks of SOI are circular Gaussian : When the SOI has the circular Gaussian pdf on the kth block, then $\overline{\kappa}_{s^k} = 1$. Hence, the CRIB (55) is infinite when there is a block where the SOI is circular Gaussian. By contrast, CRIBs (65) and (70) are finite provided that the SOI is non-Gaussian or non-circular at least on one block. In the special case when all blocks of the SOI but the kth block have circular Gaussian pdf, the CRIBs (65) and (70) say that

$$\text{CMV:}\quad \mathrm{E}[\mathtt{ISR}] \geq \frac{1}{N_b}\frac{1}{\sum_{m=1}^{M}\sigma^2_{s^m}}\times \tag{83}$$
$$\mathtt{tr}\left(\sum_{m=1}^{M}\frac{1}{\kappa_{s^m}}\mathbf{I}_{d-1} + \frac{\kappa_{s^k}}{\overline{\kappa}_{s^k}-1}\mathbf{C}_{\mathbf{z}^k}\sum_{m=1}^{M}\frac{1}{\kappa_{s^m}}\mathbf{C}^{-1}_{\mathbf{z}^m}\right),$$

$$\text{CSV:}\quad \mathrm{E}[\mathtt{ISR}] \geq \frac{1}{N_b}\frac{1}{\sum_{m=1}^{M}\sigma^2_{s^m}}\frac{\sigma^2_{s^k}}{\overline{\kappa}_{s^k}-1}\times \tag{84}$$
$$\mathtt{tr}\left(\mathbf{C}^{-1}_{\mathbf{z}^k}\sum_{m=1}^{M}\mathbf{C}_{\mathbf{z}^m}\right).$$

Consequently, for the identifiability of the CVM and CSV models, it is sufficient that the SOI is not circular Gaussian on at least one block, which is a significant advantage compared to BICE.

5) Gaussian SOI and vanishing background: When the SOI is circular Gaussian on all blocks, all the CRIBs discussed in the section are infinite, and the SOI cannot be extracted. However, we can consider a special situation where, in some block (the kth one), the SOI is close to be Gaussian and, simultaneously, the background covariance is getting close to zero. The previous special case says that the CRIBs for CMV and CSV are still finite until $\overline{\kappa}_{s^k} > 1$.

Let us consider $\overline{\kappa}_{s^m} = 1$ for $m \neq k$, and

$$\mathbf{C}_{\mathbf{z}^k} = \epsilon(\overline{\kappa}_{s^k}-1)\mathbf{T}, \tag{85}$$

where $\mathbf{T}$ is a positive definite matrix, and $\epsilon > 0$ is a constant. Now, consider $\overline{\kappa}_{s^k} \to 1$ while $\sigma^2_k = 1$, which means that the SOI is becoming circular Gaussian on the kth block while its variance is constant. Eq. (85) says that the background is vanishing proportionally to the "gaussianity" of the SOI (expressed through the proximity of $\overline{\kappa}_{s^k}$ to one).

The reader can verify, that, in that special case, the CRIB of CSV (70) is becoming infinite while that for CMV (65) approaches a finite value. This can be explained through the fact that the SOI, in the kth block, is observed without noise when $\mathbf{C}_{\mathbf{z}^k} = \mathbf{0}$. Therefore, its corresponding mixing vector can be identified through finding the principal component in that block, which is sufficient for the identifiability of the CMV model although the SOI is (almost) circular Gaussian in all blocks.

VI. Experimental Validation

In the following numerical simulations, we compare the theoretical bounds with empirical mean ISR achieved by selected ICA/ICE algorithms. Here, we have to cope with the permutation ambiguity causing that a given algorithm need not converge to the desired SOI. In case of BSE/ICE algorithms, the convergence is arranged through their proper initialization. For ICA methods, the SOI is identified as the separated signal with the lowest ISR. Since the algorithms do not converge to the right SOI in some runs, the trimmed mean of ISR is computed instead of the mean, by discarding 10% of the lowest and greatest values. Therefore, the reader should keep in mind that the empirical results can be slightly biased.

A. Determined Mixing Model

1) Gaussian background: Here, the CRIB given by (49) assuming circular Gaussian background is compared with the empirical ISR achieved by four methods. First, non-circular FastICA (NC-FastICA) from [52], is an ICA algorithm designed particularly for signals belonging to the complex Generalized Gaussian Distribution (GGD) family [9], which involves also non-circular signals. Second, OGICE (Orthogonally Constrained ICE) from [53] is an ICE algorithm derived based on maximum likelihood principle. Third, Natural Gradient (NG) is a basic ICA algorithm from [54]. In OGICE, the background is modeled as circular Gaussian, therefore, this method can asymptotically attain the CRIB in this experiment provided that the true score function of the SOI is used as the internal nonlinear function. Fourth, the RobustICA algorithm from [55] is a BSE method based on the optimization of the kurtosis contrast function. It is valid for real as well as complex-valued sources, with circular and non-circular distributions.

In a trial, $d = 5$ independent complex-valued signals are generated. The target signal is drawn from the complex-valued GGD with zero mean, unit variance, shape parameter $\alpha \in$

$(0, +\infty)$, and a circularity coefficient $\gamma \in [0, 1]$. The pdf is given by [5]

$$p(s, s^*) = \frac{\alpha\rho \exp\left(-\left[\frac{\rho/2}{\gamma^2-1}\left(\gamma s^2 + \gamma(s^*)^2 - 2ss^*\right)\right]^\alpha\right)}{\pi\Gamma(1/\alpha)(1-\gamma^2)^{\frac{1}{2}}}, \tag{86}$$

where $\rho = \frac{\Gamma(2/\alpha)}{\Gamma(1/\alpha)}$, and $\Gamma(\cdot)$ is the Gamma function. The other (background) signals are circular Gaussian, which corresponds to $\alpha = 1$ and $\gamma = 0$ in (86). All signals are mixed by a random mixing matrix $\mathbf{A}$ with elements drawn from $\mathcal{CN}(0, 1)$.

OGICE is initialized by a randomly perturbed first column of $\mathbf{A}$, Natural Gradient is initialized by the randomly perturbed mixing matrix $\mathbf{A}$, RobustICA is initialized by the randomly perturbed demixing matrix $\mathbf{W}$, while the initialization of NC-FastICA is random in full. In OGICE and NG, the nonlinearity is the same as the true score function corresponding to (86), that is,

$$\psi(s, s^*) = \frac{2\alpha(\rho/2)^\alpha}{(\gamma^2-1)^\alpha}\left(\gamma s^2 + \gamma(s^*)^2 - 2ss^*\right)^{\alpha-1}(\gamma s - s^*). \tag{87}$$

It is worth noting that the true score function is not known in a fully blind situation, where ψ must be replaced by a suitable nonlinearity; see, e.g., [56], [57]. It can be shown that [5]

$$\overline{\kappa} = \mathrm{E}\left[|\psi(s)|^2\right] = \frac{\alpha^2\Gamma(2/\alpha)}{(1-\gamma^2)\Gamma^2(1/\alpha)}. \tag{88}$$

Finally, note that NC-FastICA is endowed by the nonlinearity proposed in [52], the accuracy seems to be closest to the CRIB for $\alpha = 0.8$. RobustICA is using kurtosis [55] and seems to be efficient for sub-gaussian sources $\alpha > 1$.

Figs. 1–3 show average ISR achieved by the algorithms in 100 trials, respectively, for varying N, α, and γ. The average ISRs achieved by OGICE are very close to the bound (49), which is in a good agreement with the theory. Also RobustICA attains the bound for greater number of samples N. The performance of NC-FastICA appears to be limited, which can be explained by the nonlinearity used. The performance of NG is also limited due to convergence issues, especially, in cases of sub-gaussian ($\alpha > 1$) SOI.

In Fig. 2, the ISR for sub-gaussian ($\alpha > 1$) and super-Gaussian ($\alpha < 1$) SOI is shown. For $\alpha = 1$, all signals, including the SOI, are circular Gaussian, in which case the mixing coefficients are not identifiable. Here, the algorithm's empirical ISRs drop down to 0 dB (a highly biased value as ISR higher than 0 dB is evaluated as convergence to a different source), which should be interpreted as failings in finding the SOI.

In Fig. 3, the non-circular Gaussian SOI with varying circularity is considered. The ISR achieved by OGICE approaches the CRIB, which confirms the fact that a non-circular Gaussian signal can be extracted from the other Gaussian signals when their circularity coefficient is different. This condition becomes violated as γ approaches 0, which corresponds with the decaying ISR. NC-FastICA and RobustICA are designed to be robust to circularity changes, however, for Gaussian sources they do not benefit from non-circularity. Therefore, their performance does not show any dependence on γ and is the same as for the circular Gaussian SOI [52].

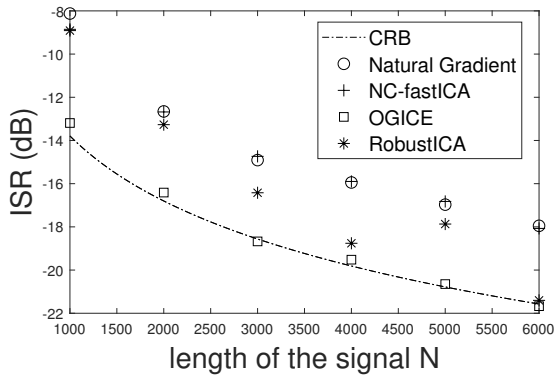

Fig. 1. Average ISR for $d = 5$, $\alpha = 2$, and varying N.

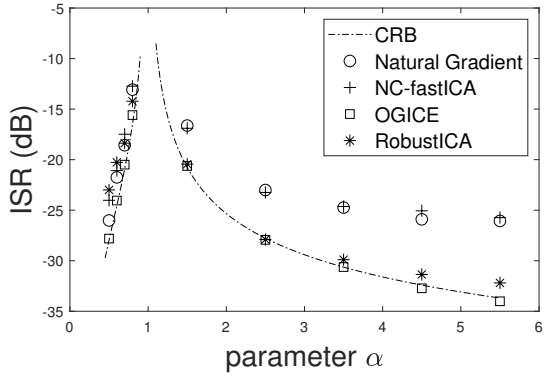

Fig. 2. Average ISR for $d = 5$, $N = 2500$ and varying α.

2) Non-Gaussian background: As shown in Section III-G, there is a coincidence between the CRIBs for ICA and ICE when, in ICE, the non-Gaussianity of background is taken into account. In this section, we simulate the case mentioned at the end of that section, that is, when background signals are dependent (a transformation decomposing them into independent components as assumed in ICA need not to exist). The theoretical CRIB for this simulation is given by (46).

In a trial, $d = 4$ real-valued signals are generated. The

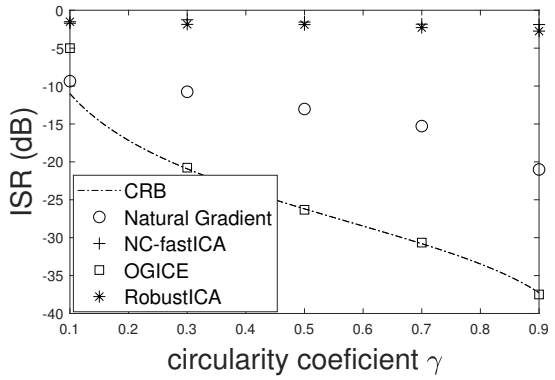

Fig. 3. Average ISR for $d = 5$, $N = 2500$, $\alpha = 1$ (Gaussian SOI) and varying circularity coefficient γ.

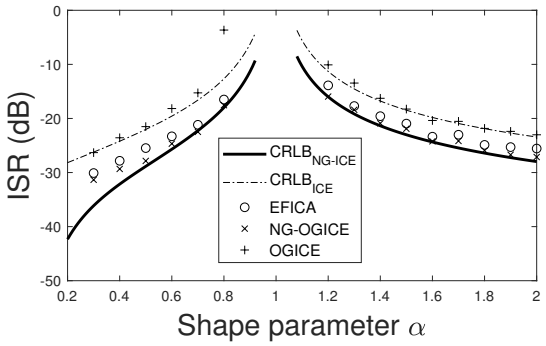

Fig. 4. Average ISR for non-Gaussian background when pdfs of all signals are varying with respect to α.

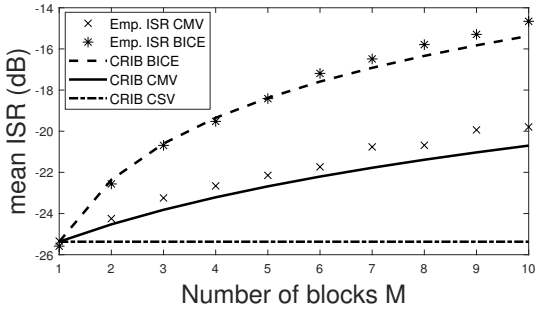

Fig. 5. Average ISR for CMV mixing model when $d = 5$, $N = 5040$, and varying number of blocks M.

background is drawn according to the joint pdf given by

$$p(z^1, \ldots, z^{d-1}) \propto \exp\left(-\left(\lambda \sum_{i=1}^{d-1} |z^i|^2\right)^{\alpha}\right) \qquad (89)$$

where $\lambda > 0$, and $\alpha \neq 1$ (for $\alpha = 1$, the pdf is Gaussian). To scale the marginal pdfs of background signals to the unit variance, we put $\lambda = \frac{\Gamma\left(\frac{5}{2\alpha}\right)}{3\Gamma\left(\frac{3}{2\alpha}\right)}$. Then, it holds that

$$(\boldsymbol{\kappa}_{\mathbf{z}})_{kk} = \frac{4}{3}\lambda\alpha^2 \frac{\Gamma(2 + \frac{1}{2\alpha})}{\Gamma(\frac{3}{2\alpha})}. \qquad (90)$$

The SOI is drawn from the real-valued GGD family [58] with zero mean, unit variance and a shape parameter $\tilde{\alpha}$, where $\tilde{\alpha} = \alpha + 1$. Note that for the real-valued GGD, the SOI is Gaussian when $\alpha = 1$; see Appendix B in [58].

We compare three algorithms with the CRIB given by (46): OGICE [53], EFICA [58], and NG-OGICE [50]. OGICE is designed for ICE with Gaussian background, where the CRIB is given by (49) (which we show as well for the sake of completeness). EFICA is an asymptotically efficient ICA algorithm provided that all original signals are drawn from the real-valued GGD. NG-OGICE is an ICE method considering the non-Gaussianity of background, in which the true multivariate score function of background must be known to achieve the optimum performance.

In Fig. 4, the ISRs averaged over 100 trials achieved by OGICE, EFICA and NG-OGICE are compared. The bound (46) is denoted by $\text{CRIB}_{\text{NG}-\text{ICE}}$ and the one for the Gaussian background (49) is denoted by CRIB_{ICE}. The results show that the mean ISRs by OGICE are close to the bound given by (49) (which is in a good agreement with the results of asymptotic performance analyses (6) [45]). The results by EFICA and NG-OGICE are closer to (46). NG-OGICE is even slightly more accurate than EFICA, which is caused by a more accurate modeling of the background's pdf.

For $\alpha = 1$, all signals are Gaussian, which means that the SOI cannot be separated from the background. With increasing non-Gaussianity of the mixture, which means increasing distance from $\alpha = 1$, the separation accuracy gets better.

B. Piecewise Determined Mixtures with Circular Gaussian Background

To validate the bounds for CMV and CSV, both are compared with empirical results achieved by block-wise versions of OGICE introduced in [12]. The methods will be jointly referred to as BOGICE (in [12], $\text{BOGICE}_{\mathbf{a}}$ is the variant for CMV while $\text{BOGICE}_{\mathbf{w}}$ is for CSV). It should be noted that no other methods for CMV/CSV currently exist in the literature to our best knowledge. For completeness, the BICE method is compared to BOGICE in both cases.

In experiments here, we consider two statistical models of signals: The SOI is either i.i.d. non-Gaussian over all blocks or i.i.d. within blocks with the same distribution but varying variance over blocks. The background is assumed circular Gaussian i.i.d. with unit variance in all blocks in both cases.

In trials, $d = 5$ independent complex-valued signals are generated. The SOI is drawn from a circular complex GGD with zero mean, unit variance, $\alpha = 2$. The other signals are circular Gaussian, which corresponds to $\alpha = 1$. The nonlinearity is given by the true score function. M blocks of the same length are considered. Each block is mixed by a random mixing matrix. The mixing matrices obey the mixing models CMV or CSV, respectively.

The empirical ISRs achieved by BOGICE and BICE are compared with the CRIB corresponding to the mixing model used in the given simulation and with that of BICE. For completeness, we also show the hypothetical CRIB of the alternative piecewise mixing model that would be valid when the mixing matrix obeyed the model and the SOI had the same statistical properties. Nevertheless, it should be kept in mind that CMV and CSV are incompatible unless all the mixing parameters related to the SOI are constant over the blocks (which is not the case of the experiments here). The comparison of the models thus has to be done after considerable deliberation.

1) An i.i.d. SOI: Fig. 5 corresponds to the simulation considering the CMV model for varying number of blocks, that is, $M = 1, 2, 5, 10$. It shows the mean ISR achieved by BOGICE averaged over 500 trials and the CRIB given by (74) (CMV) and, for comparison, also the CRIBs (73) (BICE) and (75) (CSV). Similar simulation was done with the CSV model; the results are shown in Fig. 6. As can be seen from (73), (74) and (75), bounds for BICE and CMV depend on the number

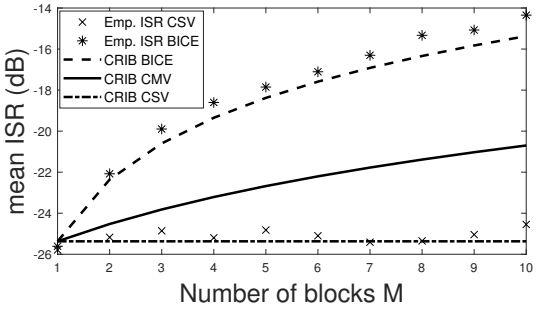

Fig. 6. Average ISR for CSV mixing model when $d = 5$, $N = 5040$, and varying number of blocks M.

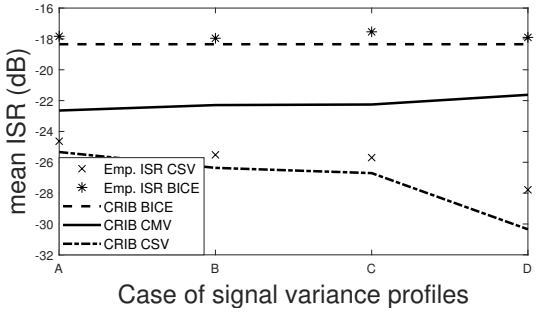

Fig. 8. Average ISR for CSV mixing model when $d = 5$, $N = 5000$, and varying σ_{s^m} over blocks.

of blocks, but the CRIB for CSV does not. Hence, the CSV curve is flat as predicted by our theoretical analysis.

Figs. 5 and 6 show the coincidence between the empirical results by the variants of BOGICE and the CRIBs corresponding to the mixing model of the given simulation. The performances of the methods follow the same dependence on the number of blocks M as these CRIBs. The results also show that BOGICE takes the advantage of the special mixing model CMV/CSV compared to BICE, as its mean ISR is lower that the CRIB (73), unless $M = 1$ where all mixing models coincide.

2) SOI with varying variance: In this special case, the SOI with the same pdf but varying variance over blocks is assumed. In a trial, $M = 5$ blocks and four different settings of SOI's variances are considered: Specifically, type A is $\sigma^2_{s^m} = 1$ for $m = 1, \ldots, 5$, type B corresponds to $\sigma^2_{s^1} = \sigma^2_{s^2} = 1$, $\sigma^2_{s^3} = 2$, $\sigma^2_{s^4} = \sigma^2_{s^5} = 3$, type C corresponds to $\sigma^2_{s^m} = m$, and type D is for $\sigma^2_{s^m} = m^2$, $m = 1, \ldots, 5$.

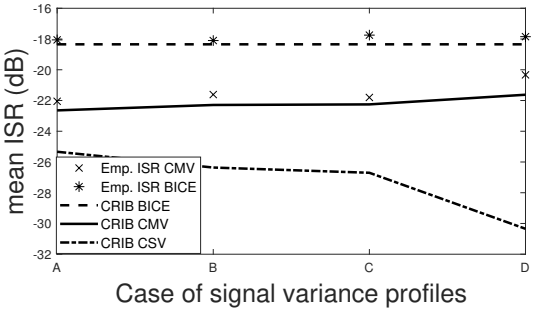

Fig. 7. Average ISR for CMV mixing model when $d = 5$, $N = 5000$, and varying σ_{s^m} over blocks.

We have analyzed in Section V-3 that the nonstationarity of the SOI improves the separation accuracy under the CSV mixing model but typically worsens the accuracy under the CMV model. The results in Figures 7 and 8 confirm this property, although the drop in performance in case of the CMV model due to the SOI's nonstationarity is not that substantial, in this experiment.

VII. Conclusions

The present contribution has computed the CRLB-induced bounds for the ISR in the ICE model, i.e., in BSE under statistically independent sources. The developed CRLBs are valid for both circular and non-circular sources, and include the scenarios of determined mixing and piecewise (block) determined mixing. The derived bounds depend on the target signal distribution and on the length of data, and they coincide with that for ICA when all but the target signals are circular Gaussian (shown for the standard mixing model). A variety of experimental results confirm the validity of the derived CRIBs. In particular, the CRIB was shown to be attainable by the OGICE algorithm when the target signal is non-Gaussian or non-circular Gaussian, under the assumption that the true nonlinearities (score functions) defining the pdf of the SOI are known in advance. Similarly, the variants of BOGICE can attain the CRIBs valid for the CMV and CSV mixing models.

References

[1] 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.

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

[3] 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.

[4] E. Doron, A. Yeredor, and P. Tichavský, "Cramér-Rao-induced bound for blind separation of stationary parametric gaussian sources," *IEEE Signal Processing Letters*, vol. 14, no. 6, pp. 417–420, June 2007.

[5] 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.

[6] 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.

[7] M. Taseska and E. A. P. Habets, "Blind source separation of moving sources using sparsity-based source detection and tracking," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 3, pp. 657–670, 2018.

[8] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," in *IEEE Trans. Information Theory*, vol. 52, March 2006, pp. 1017–1029.

[9] M. Novey, T. Adali, and A. Roy, "A complex generalized gaussian distribution–characterization, generation, and estimation," in *IEEE Trans. Signal Processing*, vol. 58, March 2010, pp. 1427–1433.

[10] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[11] 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.

[12] Z. Koldovský, J. Málek, and J. Janský, "Extraction of independent vector component from underdetermined mixtures through block-wise determined modeling," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 7903–7907.

[13] S. Bhinge, R. Mowakeaa, V. D. Calhoun, and T. Adalı, "Extraction of time-varying spatiotemporal networks using parameter-tuned constrained iva," *IEEE Transactions on Medical Imaging*, vol. 38, no. 7, pp. 1715–1725, July 2019.

[14] P. J. Huber, "Projection pursuit," *Ann. Statist.*, vol. 13, no. 2, pp. 435–475, June 1985.

[15] J. Herault and C. Jutten, "Space or time adaptive signal processing by neural network models," in *AIP Conference Proceedings 151 on Neural Networks for Computing*. Woodbury, NY, USA: American Institute of Physics Inc., 1987, pp. 206–211.

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

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

[18] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. John Wiley & Sons, 2002.

[19] T. Cover and J. Thomas, *Elements of Information Theory*. Wiley, 2006.

[20] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 530–538, Sep. 2004.

[21] T. Adalı, Y. Levin-Schwartz, and V. D. Calhoun, "Multimodal data fusion using source separation: Two effective models based on ica and iva and their properties," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1478–1493, Sep. 2015.

[22] X. Chen, Z. J. Wang, and M. McKeown, "Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods," *IEEE Signal Processing Magazine*, vol. 33, no. 3, pp. 86–107, May 2016.

[23] V. Kautsk, P. Tichavsk, Z. Koldovsk, and T. Adal, "Performance bounds for complex-valued independent vector analysis," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4258–4267, 2020.

[24] I. Lee, T. Kim, and T.-W. Lee, "Fast fixed-point independent vector analysis algorithms for convolutive blind source separation," *Signal Processing*, vol. 87, no. 8, pp. 1859–1871, 2007.

[25] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 9, pp. 1626–1641, 2016.

[26] D. Kitamura, S. Mogami, Y. Mitsui, N. Takamune, H. Saruwatari, N. Ono, Y. Takahashi, and K. Kondo, "Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation," *EURASIP Journal on Advances in Signal Processing*, vol. 2018, no. 1, p. 28, May 2018.

[27] D.-T. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1837–1848, Sep 2001.

[28] P. Tichavský and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 878–891, March 2009.

[29] A. Yeredor, "Blind separation of gaussian sources with general covariance structures: Bounds and optimal estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5057–5068, Oct 2010.

[30] G. Chabriel, M. Kleinsteuber, E. Moreau, H. Shen, P. Tichavský, and A. Yeredor, "Joint matrices decompositions and blind source separation: A survey of methods, identification, and applications," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 34–43, May 2014.

[31] 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.

[32] M. Anderson, T. Adali, and X. Li, "Joint blind source separation with multivariate gaussian model: Algorithms and performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 4, pp. 1672–1683, April 2012.

[33] 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.

[34] L. D. Lathauwer and J. Castaing, "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization," *IEEE Transactions on Signal Processing*, vol. 56, no. 3, pp. 1096–1105, March 2008.

[35] A. Ferreol, L. Albera, and P. Chevalier, "Fourth-order blind identification of underdetermined mixtures of sources (fobium)," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1640–1653, May 2005.

[36] P. Comon and M. Rajih, "Blind identification of under-determined mixtures based on the characteristic function," *Signal Processing*, vol. 86, no. 9, pp. 2271 – 2281, 2006, special Section: Signal Processing in UWB Communications.

[37] P. Tichavský and Z. Koldovský, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1037–1047, March 2011.

[38] Z. Koldovský, P. Tichavský, A. H. Phan, and A. Cichocki, "A two-stage mmse beamformer for underdetermined signal separation," *IEEE Signal Processing Letters*, vol. 20, no. 12, pp. 1227–1230, Dec 2013.

[39] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.

[40] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, "Underdetermined blind separation of convolutive mixtures of speech by combining time-frequency masks and ICA," in *Proc. of the 18th International Congress on Acoustics (ICA 2004)*, vol. I, Apr. 2004, pp. 321–324.

[41] F. Abrard and Y. Deville, "A timefrequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Signal Processing*, vol. 85, no. 7, pp. 1389 – 1403, 2005.

[42] Y. Li, S. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Transactions on Signal Processing*, vol. 54, no. 2, pp. 423–437, Feb 2006.

[43] B. Liu, V. G. Reju, and A. W. H. Khong, "A linear source recovery method for underdetermined mixtures of uncorrelated ar-model signals without sparseness," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4947–4958, Oct 2014.

[44] 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.

[45] 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.

[46] 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.

[47] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. Dover Publications, 2008.

[48] 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.

[49] J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec 1996.

[50] Z. Koldovský, P. Tichavský, and N. Ono, "Orthogonally-constrained extraction of independent non-gaussian component from non-gaussian background without ICA," in *Latent Variable Analysis and Signal Separation*, Y. Deville, S. Gannot, R. Mason, M. D. Plumbley, and D. Ward, Eds. Cham: Springer International Publishing, 2018, pp. 161–170.

[51] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, and S. Hosseini, "Blind separation of piecewise stationary non-gaussian sources," *Signal Processing*, vol. 89, no. 12, pp. 2570 – 2584, 2009, special Section: Visual Information Analysis for Security.

[52] M. Novey and T. Adali, "On extending the complex fastica algorithm to noncircular sources," in *IEEE Trans. Signal Processing*, vol. 56, May 2008, pp. 2148–2154.

[53] 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.

[54] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Proceedings of Neural Information Processing Systems*, 1996, pp. 757–763.

[55] V. Zarzoso and P. Comon, "Robust independent component analysis by iterative maximization of the kurtosis contrast with algebraic optimal step size," *IEEE Transactions on Neural Networks*, vol. 21, no. 2, pp. 248–261, Feb 2010.

[56] D. T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Transactions on Signal Processing*, vol. 45, no. 7, pp. 1712–1725, Jul 1997.

[57] P. Tichavský, Z. Koldovský, and E. Oja, "Speed and accuracy enhancement of linear ica techniques using rational nonlinear functions," in *Independent Component Analysis and Signal Separation*, M. E. Davies, C. J. James, S. A. Abdallah, and M. D. Plumbley, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 285–292.

[58] 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.