Double Nonstationarity: Blind Extraction of Independent Nonstationary Vector/Component from Nonstationary Mixtures — Performance Analysis

Václav Kautský, *Member, IEEE*, Zbyněk Koldovský, *Senior Member, IEEE*, Tülay Adalı, *Fellow, IEEE*

***Abstract*—Non-Gaussianity and non-stationarity are strong features on the basis of which blind source extraction (BSE) becomes a powerful signal processing tool. The recently proposed double nonstationarity model exploits both properties in the mixing and source models, which significantly broadens the class of identifiable signals. In this article, Cramér-Rao and performance analyses are presented, including the complex-valued case, non-circularity, joint extraction, and non-stationary mixing useful for moving source extraction. Besides identifiability conditions and achievable extraction accuracy, the results reveal the influence of a source model misspecification. Of particular interest is the case when the source of interest is Gaussian, which is not identifiable without taking into account source non-stationarity. The validity of the analyses is experimentally confirmed and compared with the empirical performance of the FastDIVA algorithm. It is shown that the closed-form expression obtained from the analysis can be used as information about the achieved interference-to-signal ratio without knowing the ground-truth signals.**

I. Introduction

EXTRACTING the desired source, referred to as source of interest (SOI), from a multi-channel mixture of signals is a fundamental task in signal processing. This paper deals with BSE where no prior information about the SOI is required, but only information about its statistics is leveraged. Similar to independent component analysis (ICA) [1], the blind extraction framework presented here is based on the assumption that the SOI is statistically independent of the other source signals in the mixture. We consider a recent advanced model that involves two types of nonstationarity: the nonstationarity of the mixing process and the nonstationarity of the SOI, which is referred to as *double nonstationarity*. It was introduced by Koldovský et al. [2] together with a fast algorithm which is capable of extracting a Gaussian SOI, a task that has not been possible without admitting the nonstationarity of the SOI. In addition, the algorithm involves the extension to joint blind source extraction (jBSE) where several mixtures[1] are processed jointly.

This work is focused on theoretical and performance analyses of the double nonstationarity model and of the algorithm from [2]. The analyses bring new findings about the achievable accuracy of extraction based on the model, identifiability conditions, and about the statistical optimality of the algorithm. The algorithm analysis provides insight into the problem of mismatch between the true and model distribution of the SOI, where the latter must be chosen since the true distribution is not known. Of particular interest is the Gaussian case: We address the problem of how the extraction accuracy is affected in jBSE when the model covariance matrix of the multivariate Gaussian SOI differs from the true covariance matrix.

The idea of BSE already appeared in the past [3]. Under the framework of ICA, it is realized, e.g., by the one-unit version of the famous FastICA algorithm [4], [5]. The approach was revised in [6] on the basis of a reduced parameterization of the mixing matrix, and it was proved to be closely related or even equivalent to the non-Gaussianity-based approaches of the 90s [5], [7], [8][2]. The approach in [6] was referred to as independent component extraction (ICE) when a single mixture was addressed, and independent vector extraction (IVE) for the extended problem of jBSE. Then, ICE and IVE have been extended for nonstationary (dynamic) mixing in [9], where the mixing model is semi-time-variant as it assumes a constant separating vector (CSV). The CSV model allows changes to the SOI mixing parameters; for example, it has been proven useful in moving speaker extraction [10]; its theoretical merits have been shown in [11]. In [9], FastICA has been extended for CSV, which brought a powerful method for the nonstationary mixing model. The double nonstationarity-based model in [2] brings further substantial improvement to this algorithm. The lack of analysis of the model and algorithm is a knowledge gap that this paper fills.

A. Background

Blind source separation (BSS) aims at recovering all sources from an observed mixture without prior information. The BSE could be seen as a subtask of BSS where one particular

V. Kautský is with the Acoustic Signal Analysis and Processing Group, Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, Czech Republic and with the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Czech Republic. (email: kautsvac@fjfi.cvut.cz).

Z. Koldovský is with the Acoustic Signal Analysis and Processing Group, Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, Czech Republic. (email: zbynek.koldovsky@tul.cz).

T. Adalı is with the Department of Computer Science and Electrical Engineering, University of Maryland, Baltimore County, Baltimore, MD 21250, USA. (email: adali@umbc.edu).

This work was supported by the Czech Science Foundation through Project No. 20-17720S, by the Department of the Navy, Office of Naval Research Global, through Projects No. N62909-19-1-2105 and No. N62909-23-1-2084, and by the US National Science Foundation through Project No. 2316420.

[1] In data fusion literature but also in [2], the mixtures have been referred to as *data-sets*. In this paper, we shall use the term "mixtures".

[2] For example, the algorithm in [9] is an extension of the one-unit FastICA from [5]; also, its performance analysis in [9] shows the same statistical accuracy as the algorithm of [4].

source is of interest. ICA provides a powerful BSS approach assuming that the sources are statistically independent [12]. Simultaneous separation of multiple mixtures is known as joint blind source separation (jBSS) [13] where the extension of ICA for jBSS is independent vector analysis (IVA) [14], and more recently also independent low-rank matrix analysis (ILRMA) [15] where joint models of sources are sought through nonnegative matrix factorization (NMF) [16]. The BSE and jBSE alternatives to ICA resp. IVA is the ICE resp. IVE [6]. For BSE/jBSE, it is possible to consider new mixing models as recently demonstrated, e.g., in [2], [9], [17].

Blind methods address the typical situation where no ground-truth information about the mixture and sources is available. It is therefore important to provide answers to the following questions: (i) what are the conditions under which the parameters of the given model can be consistently estimated (identification conditions)? and (ii) what is the best achievable estimation accuracy, and what are the limitations caused by mismatched mixing or source models? To address these questions, the Cramér-Rao lower bound (CRLB) provides a powerful analytic tool.

Regarding ICA and its extensions, the CRLB has been studied for various signal models. For real-valued non-Gaussian independent and identically distributed (i.i.d.) sequences (often called higher-order statistics (HOS)-based ICA), the CRLB shows that all but one source must be non-Gaussian for the full identifiability of the model (e.g., [18], [19]). When multiple signal properties are taken into account, e.g., HOS and sample dependence, the class of signals that can be identified includes multiple Gaussians [20], [21], and with the addition of noncircularity for complex-valued signals, this includes i.i.d. Gaussians as long as they have distinct circularity coefficients [22], [23]. In HOS-based ICE, only the SOI must be non-Gaussian [24] while the other source signals (which are now referred to as *background*) can be Gaussian; a more detailed analysis in [11] considers also ICE with a non-Gaussian background. For HOS-based IVA, the CRLB uncovers how dependencies among sources from different mixtures extend the class of identifiable sources and improve the accuracy [20], [25]; the CRLB for the complex-valued HOS-based IVA was analyzed in [26] where also the influence of non-circularity of the SOI was taken into account; for the CRLB related to the corresponding IVE model, see [27].

Very useful information is gained by comparing the CRLB with the numerical or analytical performance of a given algorithm. Of particular interest is the performance loss due to model misspecification [28]. It uncovers how the performance is influenced by properties of signals such as non-Gaussianity, non-circularity, nonstationarity, non-whiteness, and correlations and higher-order dependencies in the case of jBSS/jBSE [20], [29]. HOS-based ICA algorithms were analyzed and compared with the corresponding CRLB, e.g., in [19], [30], [31].

Methods based on second-order statistics (SOS) are attractive as they provide efficient solutions, and beyond Gaussian source models, see e.g., [32]. There, the identifiability can be ensured by allowing for sources' sample dependence or nonstationarity [33]. Nonstationarity appears to be a strong signal property, characteristic of, e.g., speech or some biomedical signals [34]. An analysis of methods of this kind was done, e.g., in [35]. jBSS with Gaussian sources was first considered in [36]; IVA with a Gaussian model was analyzed in [37], [38]. Gaussian BSE and jBSE were analyzed in [39] and [40], respectively. There are also studies dedicated to combined source models, e.g., ICA based on non-Gaussianity and nonstationarity [41], BSS based on Gaussianity and non-whiteness [32], [35]; as well as HOS and sample dependence [42], and noncircularity, HOS and sample dependence [22].

B. Contribution

1) CRLB for double nonstationarity: First, we analyze the double nonstationarity-based jBSE model by computing the corresponding CRLB. Here, the mixing parameters are nonstationary as in CSV (a diversity motivated by time-varying mixtures of moving sources), and the source model of the SOI allows for (non-)Gaussianity as well as nonstationarity. The CRLB shows that —as expected— both types of nonstationarity contribute to the achievable mean interference-to-signal ratio (ISR) of the extracted sources(s). Moreover, it shows that the SOI's nonstationarity enables the model identification even when its distribution is Gaussian, which is not identifiable for the previous HOS-based model in [9].

2) Performance analysis: Second, we analyze the stationary point of the contrast function that is derived through a modification of the likelihood function. This leads to the performance analysis of the algorithms in [2]. The analysis reveals conditions under which the algorithms are asymptotically efficient. The analysis also uncovers performance loss due to a misspecified source model, which is particularly interesting for the complex-valued IVE (jBSE) where the SOI is multivariate Gaussian with correlations and pseudo-correlations (non-circularity) across the mixtures. Another practical outcome is the ISR post-estimation. We show that the closed-form expression for the mean ISR can be used as information about the achieved ISR, which is otherwise unknown without knowledge of the ground-truth signals.

The article is organized as follows. In Section II, the mathematical formulation of the problems addressed by this article is presented. Sections III and IV are, respectively, devoted to the statistical model and to the contrast function for parameter estimation. Section V introduces the mean ISR criterion that is used in the CRLB analysis in Section VI and for the performance analysis in Section VII. A discussion follows this in Section VIII and the experimental validation in Section IX. Section X concludes the article.

II. Problem Formulation

A. Sample partition

We consider K mixtures, each containing N D-dimensional observations of latent, unknown, signals. Within each mixture, we divide the N samples into T non-overlapping intervals of the same[3] number of samples N_b, called blocks, and each

[3]The analyses presented in this paper can be easily extended to the case where different blocks and sub-blocks have different numbers of samples.

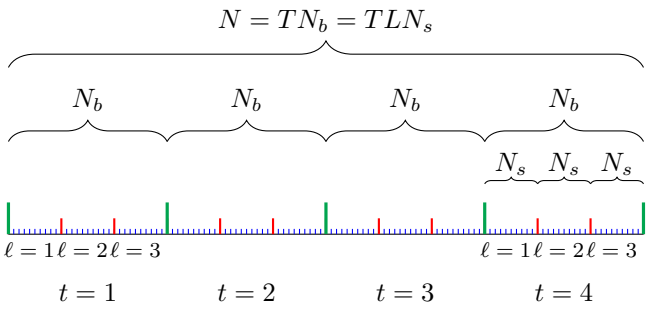

Fig. 1. Illustration of the block-partition of the samples of signals. In this example, the number of samples per sub-block is $N_s = 10$, the number of blocks is $T = 4$, and the number of sub-blocks per block is $L = 3$. Therefore, the number of samples per block is $N_b = LN_s = 30$, and the number of samples is $N = TN_b = TLN_s = 4 \cdot 3 \cdot 10 = 120$.

TABLE I
GLOSSARY OF MAIN VARIABLES AND PARAMETERS

Variable	Description	Range
K	number of mixtures	$k = 1, \ldots, K$
D	number of sensors	not used
N	number of samples per mixture	not used
T	number of blocks (partitions of N)	$t = 1, \ldots, T$
N_b	number of samples per block	not used
L	number of sub-blocks per block	$\ell = 1, \ldots, L$
N_s	number of samples per sub-block	$n = 1, \ldots, N_s$

block is further partitioned into L sub-blocks with N_s samples each. Hence, $N = T \cdot N_b$, $N_b = L \cdot N_s$, and $N = T \cdot L \cdot N_s$. The purpose of this partitioning is to allow for changes (non-stationarity) in the mixing model across blocks and even finer changes in the source model across sub-blocks. While K is given as the "third" dimension of the input data, T and L should be chosen according to the application. Throughout this paper, we shall denote the index of mixture, block, and sub-block, respectively, by $k = 1, \ldots, K$, $t = 1, \ldots, T$, and $\ell = 1, \ldots, L$. Figure 1 illustrates the block-partition of the samples of signals. Table I summarizes the main variables and parameters used in this paper.

B. Linear mixture model

In the kth mixture, tth block, and ℓth sub-block, the nth sample of the observed signals, $n = 1, \ldots, N_s$, is represented by the observations vector $\mathbf{x}_{t,\ell,n}^{[k]} \in \mathbb{C}^{D\times 1}$, where D is the number of sensors in each mixture. The observations at each sub-block are assumed to satisfy the linear instantaneous mixing model:

$$\mathbf{x}_{t,\ell,n}^{[k]} = \mathbf{A}_t^{[k]} \mathbf{u}_{t,\ell,n}^{[k]}\,, \tag{1}$$

where $\mathbf{A}_t^{[k]} \in \mathbb{C}^{D\times D}$ is the *mixing matrix* at mixture k and block t, and $\mathbf{u}_{t,\ell,n}^{[k]} \in \mathbb{C}^{D\times 1}$ is the vector consisting of all sources in the mixture. The first entry of

$$\mathbf{u}_{t,\ell,n}^{[k]} = \begin{bmatrix} s_{t,\ell,n}^{[k]} & (\mathbf{z}_{t,\ell,n}^{[k]})^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{D\times 1}$$

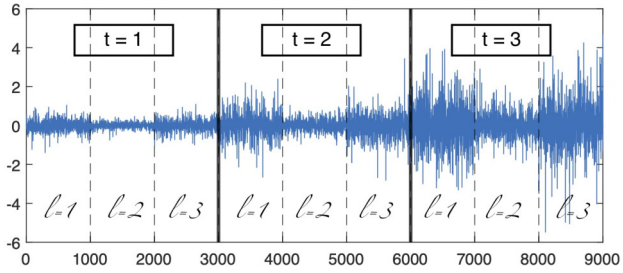

Fig. 2. An example of a block-stationary signal consisting of $N = 9000$ samples, divided into $T = 3$ blocks, each containing $L = 3$ sub-blocks; the signal is i.i.d. within each sub-block.

is associated with the SOI $s_{t,\ell,n}^{[k]}$, whereas the remaining $D-1$ entries $\mathbf{z}_{t,\ell,n}^{[k]}$ are associated with *background* sources (signals) that are not of interest.

C. Uncertainties

BSS problems involve inherent uncertainties and unknowns, such as the order of the separated signals (see, e.g., [43]), their signs, phases, and scales (see, e.g., [44], [45]). These uncertainties are manifested in BSE by the fact that the role of SOI can play a different (independent) signal; the uncertainty of its scale, sign, and phase arises from the fact that $s_{t,\ell,n}^{[k]}$ can be replaced in (1) by $\alpha_t^{[k]} s_{t,\ell,n}^{[k]}$ when the first column of $\mathbf{A}_t^{[k]}$ is multiplied by $1/\alpha_t^{[k]}$ where $\alpha_t^{[k]} \neq 0$, while $\mathbf{x}_{t,\ell,n}^{[k]}$ remain unchanged.

D. Piecewise stationary model

Statistical assumption #1: In this paper, we assume that the source samples $\mathbf{u}_{t,\ell,n}^{[k]}$ are well approximated by realizations of zero-mean piecewise-stationary random processes with statistically independent samples. Within each sub-block, the random vectors $\mathbf{u}_{t,\ell,n}^{[k]}$ are i.i.d.; the argument n shall therefore be omitted if not explicitly needed. Figure 2 exemplified a block-stationary signal partitioned into these blocks.

E. Extracting the SOI

Source recovery: The source vector $\mathbf{u}_{t,\ell,n}^{[k]}$ can be recovered by applying the inverse transformation $\mathbf{W}_t^{[k]} = (\mathbf{A}_t^{[k]})^{-1} \in \mathbb{C}^{D\times D}$ to the observations:

$$\mathbf{u}_{t,\ell,n}^{[k]} = \mathbf{W}_t^{[k]} \mathbf{x}_{t,\ell,n}^{[k]}\,. \tag{2}$$

As we shall show in this section, there is no need to estimate the whole matrix $\mathbf{A}_t^{[k]}$ or its inverse, if we are interested in recovering only the SOI. Instead, we can suffice with characterizing the one-dimensional subspace spanned by the first column of $\mathbf{A}_t^{[k]}$ and the $(D-1)$-dimensional subspace spanned by its last $D-1$ columns[4]. Equivalently, only the $(D-1)$-dimensional signal subspace represented by $\mathbf{z}_{t,\ell,n}^{[k]}$

[4]Note that knowing only one of these subspaces is not enough because matrix $\mathbf{A}_t^{[k]}$ is not orthogonal, in general.

should be identified. Therefore, the parameterization that we shall use has the following form [6]:

$$\mathbf{A}_t^{[k]} = \begin{pmatrix} \mathbf{a}_t^{[k]} & \mathbf{Q}_t^{[k]} \end{pmatrix} = \begin{pmatrix} \gamma_t^{[k]} & (\mathbf{h}^{[k]})^{\mathsf{H}} \\ \mathbf{g}_t^{[k]} & \frac{1}{\gamma_t^{[k]}}(\mathbf{g}_t^{[k]}(\mathbf{h}^{[k]})^{\mathsf{H}} - \mathbf{I}_{D-1}) \end{pmatrix}, \tag{3}$$

where $\mathbf{a}_t^{[k]} = \begin{bmatrix} \gamma^{[k]} & (\mathbf{g}^{[k]})^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{D\times 1}$ is associated with the SOI and is called the *mixing vector*; $\mathbf{I}_D \in \mathbb{C}^{D\times D}$ denotes the identity matrix. $\mathbf{Q}_t^{[k]} \in \mathbb{C}^{D\times(D-1)}$ collects the remaining columns of $\mathbf{A}_t^{[k]}$. Its structure is derived from the following (more intuitive) unmixing matrix model as $\mathbf{A}_t^{[k]} = (\mathbf{W}_t^{[k]})^{-1}$ where

$$\mathbf{W}_t^{[k]} = \begin{pmatrix} (\mathbf{w}^{[k]})^{\mathsf{H}} \\ \mathbf{B}_t^{[k]} \end{pmatrix} = \begin{pmatrix} (\beta^{[k]})^* & (\mathbf{h}^{[k]})^{\mathsf{H}} \\ \mathbf{g}_t^{[k]} & -\gamma_t^{[k]}\mathbf{I}_{D-1} \end{pmatrix}, \tag{4}$$

$\mathbf{w}^{[k]} = \begin{bmatrix} \beta^{[k]} & (\mathbf{h}^{[k]})^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{D\times 1}$ is called the *separating vector*, and $\beta^{[k]} \in \mathbb{C}$. Matrix $\mathbf{B}_t^{[k]} = \begin{bmatrix} \mathbf{g}_t^{[k]} & -\gamma_t^{[k]}\mathbf{I}_{D-1} \end{bmatrix} \in \mathbb{C}^{(D-1)\times D}$ is called the *blocking matrix* whose key property is that $\mathbf{B}_t^{[k]}\mathbf{a}_t^{[k]} = \mathbf{0}$. Together with an additional constraint that we impose, called the *distortionless constraint*:

$$(\mathbf{w}^{[k]})^{\mathsf{H}}\mathbf{a}_t^{[k]} = 1 \quad \text{for all } t \text{ and } k\,,$$

one can readily verify that the product of the unmixing matrix with the mixing one yields the identity matrix:

$$\mathbf{W}_t^{[k]}\mathbf{A}_t^{[k]} = \mathbf{I}_D \quad \text{for all } t \text{ and } k\,.$$

Therefore, this unmixing matrix can recover the source vector $\mathbf{u}_{t,\ell}^{[k]}$ for any t, ℓ, k:

$$\mathbf{W}_t^{[k]}\mathbf{x}_{t,\ell}^{[k]} = \mathbf{W}_t^{[k]}\mathbf{A}_t^{[k]}\mathbf{u}_{t,\ell}^{[k]} = \mathbf{u}_{t,\ell}^{[k]}\,.$$

It follows that $(\mathbf{w}^{[k]})^{\mathsf{H}}\mathbf{x}_{t,\ell}^{[k]} = s_{t,\ell}^{[k]}$ and $\mathbf{B}_t^{[k]}\mathbf{x}_{t,\ell}^{[k]} = \mathbf{z}_{t,\ell}^{[k]}$ for all k, t, and ℓ, where $\mathbf{z}_{t,\ell}^{[k]}$ span the same signal subspace as the background signals as defined in (1).

The practical implication is that in order to separate the SOI from the background (i.e., to achieve BSE or jBSE), it is enough to estimate $\mathbf{a}_t^{[k]}$ and $\mathbf{w}^{[k]}$, for each block and mixture.

F. Motivation for time-invariant separating vector

We remind that the mixing vectors $\mathbf{a}_t^{[k]}$ depend on t while the separating vector $\mathbf{w}^{[k]}$ hence its parts $\beta^{[k]}$ and $\mathbf{h}^{[k]}$ do not. This is the concept of the semi-time-varying CSV mixing model that represents one of the non-stationarities of the model, namely in the parametric mixture model. The varying mixing vectors $\mathbf{a}_t^{[k]}$ entail a changing location of the SOI, i.e., when the SOI is moving; such is the case in acoustic applications when the SOI is an audio signal of a moving speaker, for example.

The fact that $\mathbf{w}^{[k]}$ and yet its inner product with $\mathbf{a}_t^{[k]}$ do not depend on t means that it simultaneously extracts the whole region where SOI occurs during motion. The model assumes the existence of such a solution, which implies limitations and advantages. The practical applicability of the model has already been confirmed in speaker extraction and electroencephalogram data processing; see, e.g., [9], [10], [46]; usability in other technical problems is highly application dependent.

G. Statistical model for ICE and IVE

Statistical assumptions #2: We now describe additional statistical assumptions on the SOI and background signals that will allow us to extract the SOI from the observed mixtures.

The random variable that generates an SOI in each sample (t, ℓ, n) and a specific mixture k is assumed to be statistically dependent of all other SOIs at the same sample index in all other mixtures. This builds the connection between the mixtures needed for joint processing. To state this mathematically, we collect all random variables that generate such an SOI sample in all mixtures in a single vector: $\mathbf{s}_{t,\ell} = \begin{bmatrix} s_{t,\ell}^{[1]} & \cdots & s_{t,\ell}^{[K]} \end{bmatrix}^{\mathsf{T}}$, called the *vector component of the SOI* for given t, ℓ. The random variable $s_{t,\ell}^{[k]}$ shall be called the kth component of the SOI for given t, ℓ. The (joint) probability density function (pdf) of $\mathbf{s}_{t,\ell}$ for given t, ℓ shall be denoted by $p_{\mathbf{s}_{t,\ell}}(\mathbf{s}_{t,\ell})$. In general, $p_{\mathbf{s}_{t,\ell}}(\mathbf{s}_{t,\ell})$ cannot be further factorized into a product of smaller non-trivial probabilities.

Contrary to the SOI, the background signals from different mixtures are assumed to be statistically independent; the pdf of $\mathbf{z}_{t,\ell}^{[k]}$ shall be denoted by $p_{\mathbf{z}_{t,\ell}^{[k]}}(\mathbf{z}_{t,\ell}^{[k]})$.

Statistical assumptions #3: The fundamental assumption on which ICE and IVE are based is that $\mathbf{s}_{t,\ell}$ and $\mathbf{z}_{t,\ell}^{[k]}$ are statistically independent for a given sub-block t, ℓ and all mixture index k. Therefore, their joint pdf within the tth block and the ℓth sub-block is equal to:

$$p(\mathbf{s}_{t,\ell}, \mathbf{z}_{t,\ell}^{[1]}, \ldots, \mathbf{z}_{t,\ell}^{[K]}) = p_{\mathbf{s}_{t,\ell}}(\mathbf{s}_{t,\ell}) \prod_{k=1}^{K} p_{\mathbf{z}_{t,\ell}^{[k]}}(\mathbf{z}_{t,\ell}^{[k]}). \tag{5}$$

III. Likelihood

A. Probability distribution of one sample

In our analysis, it will be convenient to stack all observations from a specific sample n of a given block t and sub-block ℓ, and all mixtures, in $\mathbf{x}_{t,\ell,n} = \begin{bmatrix} (\mathbf{x}_{t,\ell,n}^{[1]})^{\mathsf{T}} & \cdots & (\mathbf{x}_{t,\ell,n}^{[K]})^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{DK\times 1}$. Similarly, we concatenate all source vectors: $\mathbf{u}_{t,\ell,n} = \begin{bmatrix} (\mathbf{u}_{t,\ell,n}^{[1]})^{\mathsf{T}} & \cdots & (\mathbf{u}_{t,\ell,n}^{[K]})^{\mathsf{T}} \end{bmatrix}^{\mathsf{T}} \in \mathbb{C}^{DK\times 1}$. Last, we build a $DK \times DK$ block-diagonal matrix, whose kth diagonal $D \times D$ block is $\mathbf{A}^{[k]}$:

$$\mathbf{A}_t = \begin{bmatrix} \mathbf{A}_t^{[1]} & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{A}_t^{[K]} \end{bmatrix}.$$

Together, the observations can be modeled as:

$$\mathbf{x}_{t,\ell,n} = \mathbf{A}_t\mathbf{u}_{t,\ell,n}\,.$$

Assuming invertibility of $\mathbf{A}_t^{[k]}$ for all t and k, we can define the matrix $\mathbf{W}_t = \mathbf{A}_t^{-1}$, whose kth diagonal block is $\mathbf{W}_t^{[k]}$. The source recovery model can be expressed as:

$$\mathbf{u}_{t,\ell,n} = \mathbf{W}_t\mathbf{x}_{t,\ell,n}\,.$$

With this compact notation, it is now convenient to write the probability distribution of $\mathbf{x}_{t,\ell,n}$:

$$p_{\mathbf{x}_{t,\ell}|\mathbf{A}_t}(\mathbf{x}_{t,\ell,n}) = \frac{1}{|\det(\mathbf{A}_t)|^2}\, p_{\mathbf{u}_{t,\ell}}(\mathbf{A}_t^{-1}\mathbf{x}_{t,\ell,n}) \tag{6a}$$

$$= |\det(\mathbf{W}_t)|^2\, p_{\mathbf{u}_{t,\ell}}(\mathbf{W}_t\mathbf{x}_{t,\ell,n}) \tag{6b}$$

$$= p_{\mathbf{s}_{t,\ell}}\left(\{(\mathbf{w}^{[k]})^{\mathsf{H}}\mathbf{x}_{t,\ell,n}\}_{k=1}^{K}\right) \prod_{k=1}^{K} p_{\mathbf{z}_{t,\ell}^{[k]}}(\mathbf{B}_t^{[k]}\mathbf{x}_{t,\ell,n})\,|\det(\mathbf{W}_t^{[k]})|^2 \tag{6c}$$

where we used the property that $\det^{-1}(\mathbf{A}) = \det(\mathbf{A}^{-1})$ for any nonsingular matrix $\mathbf{A}$. The transition to (6b) is a change of variables due to the linear observation model; since we consider complex-valued signals and parameters, the absolute value of the determinant is squared. In the last transitions, to (6c), we used the joint source distribution in (5). We also used the fact that $\mathbf{W}_t$ is block-diagonal with nonsingular diagonal blocks.

To give an overview, we list the assumptions and note their importance in Table II; some of the facts given there will follow later from the results of the analyses.

TABLE II
SUMMARY OF MODEL ASSUMPTIONS

assumption	note on importance
independence of SOI and background	principal
SOI is not stationary circular Gaussian	principal due to the identifiability of SOI
mixing model	important; a small mismatch is often possible in practice
dependencies within SOI vector component	optional, necessary for solving the permutation problem
mixing model parameters T and L	often need not match perfectly
the background is stationary circular Gaussian and uncorrelated across mixtures	model simplification only; need not be satisfied
noncircularity, non-Gaussianity and nonstationarity of SOI	at least one feature is needed

B. Log-likelihood function

Taking the logarithm of the probability distribution in (6), we obtain a function that can be regarded as the log-likelihood of the nth observed sample within the tth block and ℓth sub-block given the model parameters $\mathbf{a}_t = \{\mathbf{a}_t^{[1]}, \ldots, \mathbf{a}_t^{[K]}\}$ and $\mathbf{w} = \{\mathbf{w}^{[1]}, \ldots, \mathbf{w}^{[K]}\}$:

$$\mathcal{L}_{t,\ell}(\mathbf{a}_t, \mathbf{w} \mid \mathbf{x}_{t,\ell,n}) = \log p_{\mathbf{s}_{t,\ell}}(\{(\mathbf{w}^{[k]})^{\mathsf{H}}\mathbf{x}_{t,\ell,n}^{[k]}\}_{k=1}^{K}) + \sum_{k=1}^{K}\left(\log p_{\mathbf{z}_{t,\ell}^{[k]}}(\mathbf{B}_t^{[k]}\mathbf{x}_{t,\ell,n}^{[k]}) + \log|\det\mathbf{W}_t^{[k]}|^2\right). \tag{7}$$

According to the statistical assumptions #1–#3, the observed samples are i.i.d. within sub-blocks and independently (but not identically) distributed across the blocks and sub-blocks. Therefore, the log-likelihood for the model, divided by N for normalization, is given by

$$\mathcal{L}(\mathbf{a}, \mathbf{w}) = \frac{1}{N}\log p(\{\mathbf{x}_{t,\ell,n}\}_{t=1}^{T}{}_{\ell=1}^{L}{}_{n=1}^{N_s}) = \frac{1}{N}\sum_{\ell=1}^{L}\sum_{t=1}^{T}\sum_{n=1}^{N_s}\mathcal{L}_{t,\ell}(\mathbf{a}_t, \mathbf{w} \mid \mathbf{x}_{t,\ell,n}) = \left\langle\left\langle \widehat{\mathrm{E}}\left[\mathcal{L}_{t,\ell}(\mathbf{a}_t, \mathbf{w} \mid \mathbf{x}_{t,\ell})\right]\right\rangle_{\ell}\right\rangle_{t}, \tag{8}$$

where $\mathbf{a} = \{\mathbf{a}_1, \ldots, \mathbf{a}_T\}$, and where we introduce the sample mean operator $\widehat{\mathrm{E}}[\cdot] \triangleq \frac{1}{N_s}\sum_{n=1}^{N_s}$ (because data within sub-blocks are i.i.d.) and block and sub-block averaging operators $\langle\cdot\rangle_t \triangleq \frac{1}{T}\sum_{t=1}^{t}$ and $\langle\cdot\rangle_\ell \triangleq \frac{1}{L}\sum_{\ell=1}^{L}$, respectively.

We shall use the log-likelihood function for deriving the CRLB in Section VI. In practice, this function is not available since the pdfs $p_{\mathbf{s}_{t,\ell}}$ and $p_{\mathbf{z}_{t,\ell}^{[k]}}$ are not known. They must be replaced by suitable surrogates, and the log-likelihood function is replaced by a more convenient contrast function that is optimized by algorithms.

IV. CONTRAST FUNCTION

A. A simplified (surrogate) model for the SOI

In this paper, we choose the same type of surrogate pdf as in [9, Eq. (10)], where it was introduced due to the non-stationarity of the mixing model. We replace $p_{\mathbf{s}_{t,\ell}}(\mathbf{s}_{t,\ell})$ by[5]

$$p_{\mathbf{s}_{t,\ell}}(\mathbf{s}_{t,\ell}) \approx f\left(\left\{\frac{s_{t,\ell}^{[k]}}{\widehat{\sigma}_{t,\ell}^{[k]}}\right\}_{k=1}^{K}\right)\left(\prod_{k=1}^{K}\widehat{\sigma}_{t,\ell}^{[k]}\right)^{-2}, \tag{9}$$

where $f(\cdot)$ is a pdf of a normalized random variable (zero mean and unit variance of each scalar variable); $(\sigma_{t,\ell}^{[k]})^2$ denotes the variance of $s_{t,\ell}^{[k]}$, and $(\widehat{\sigma}_{t,\ell}^{[k]})^2$ is the sample-based estimate of $(\sigma_{t,\ell}^{[k]})^2$. This choice is motivated by the transformation theorem where a normalized model pdf is re-scaled according to the variance of the SOI within the tth block and ℓth sub-block. The pdf $f(\cdot)$ can be considered dependent on t and ℓ. However, for simplicity, we assume $f(\cdot)$ constant over the blocks and sub-blocks.

B. A simplified (surrogate) model for the background

The background signals are assumed to have the multivariate circular Gaussian distribution, although their true distribution can be non-Gaussian, especially when they involve non-Gaussian sources. This assumption brings important simplifications while the loss due to a misspecified background pdf is usually acceptable [6], [9]. Let the covariance of $\mathbf{z}_{t,\ell}^{[k]}$ be denoted by $\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}$.

C. A contrast function based on the simplified (surrogate) distributions

A contrast function is an optimization criterion such that all its global maxima (or minima) correspond to the separation

[5] In [9, Eq. (10)], there were no sub-blocks; (9) is the same as [9, Eq. (10)], the only difference being that here we add the sub-block index ℓ.

of the SOI from the background [12, Chapter 3]. The log-likelihood (8) is the "natural choice" for the contrast function; however, it must be modified since it contains unknown pdfs. These are replaced by the simplified (surrogate) models that we have just defined. Plugging them into (7) and (8), and neglecting terms that do not depend on $\mathbf{a}$ and $\mathbf{w}$, the contrast function obtains the form

$$\mathcal{C}(\mathbf{a},\mathbf{w}) = \Bigg\langle \Bigg\langle \widehat{\mathrm{E}}\Bigg[\log f\Bigg(\left\{\frac{\widehat{s}_{t,\ell}^{[k]}}{\widehat{\sigma}_{t,\ell}^{[k]}}\right\}_k\Bigg)\Bigg] - \sum_{k=1}^{K}\log(\widehat{\sigma}_{t,\ell}^{[k]})^2 - \sum_{k=1}^{K}\widehat{\mathrm{E}}\left[(\widehat{\mathbf{z}}_{t,\ell}^{[k]})^{\mathsf{H}}(\mathbf{C}_{\mathbf{z},t,\ell}^{[k]})^{-1}\widehat{\mathbf{z}}_{t,\ell}^{[k]}\right]\Bigg\rangle_{\ell} + (D-2)\sum_{k=1}^{K}\log|\gamma_t^{[k]}|^2\Bigg\rangle_t, \quad (10)$$

where

$$\widehat{s}_{t,\ell}^{[k]} = (\mathbf{w}^{[k]})^{\mathsf{H}}\mathbf{x}_{t,\ell}^{[k]} \quad \text{and} \quad \widehat{\mathbf{z}}_{t,\ell}^{[k]} = \mathbf{B}_t^{[k]}\mathbf{x}_{t,\ell}^{[k]}$$

are, respectively, the estimates of the SOI and background signals as they depend on estimates $\mathbf{a}$ and $\mathbf{w}$. In the transition from (7) to (10), we used the fact that $\det\mathbf{W}_t^{[k]} = (-1)^{D-1}(\gamma_t^{[k]})^{D-2}$ (this was shown in [6, Eq. (15)]).

Basically, (10) is similar to the contrast function in [9, Eq. (12)], except for the addition of the average over ℓ; the same holds for the likelihood functions (8) and [9, Eq. (9)].

V. Interference-to-Signal Ratio

In the following analyses, we focus on the output ISR measured relative to the SOI because this criterion appropriately evaluates extraction accuracy and has an advantageous property called equivariance, as we show below. The analyses are local in the sense that the ISR is analyzed in the vicinity of the true solution. By this, we avoid the uncertainties[6] described at the end of Section II-B.

Let $\widehat{\mathbf{w}}^{[k]}$ be the estimated separating vector for the kth mixture. For a local analysis of the ISR, let us assume that $\widehat{\mathbf{w}}^{[k]} = \mathbf{w}^{[k]} + \boldsymbol{\epsilon}^{[k]}$ where $\mathbf{w}^{[k]}$ is the true separating vector and $\boldsymbol{\epsilon}^{[k]}$ is an error vector of the asymptotic order $o(1)$. The extracted signal can be expressed as

$$\begin{aligned}\widehat{s}_{t,\ell}^{[k]} &= (\widehat{\mathbf{w}}^{[k]})^{\mathsf{H}}\mathbf{x}_{t,\ell}^{[k]} \\ &= (\widehat{\mathbf{w}}^{[k]})^{\mathsf{H}}\underbrace{\mathbf{A}_t^{[k]}\mathbf{W}_t^{[k]}}_{\mathbf{I}_D}\mathbf{x}_{t,\ell}^{[k]} = (\mathbf{e}_1 + (\mathbf{A}_t^{[k]})^{\mathsf{H}}\boldsymbol{\epsilon}^{[k]})^{\mathsf{H}}\mathbf{u}_{t,\ell}^{[k]} \\ &= (1+\epsilon^{[k]})s_{t,\ell}^{[k]} + (\mathbf{q}^{[k]})^{\mathsf{H}}\mathbf{z}_{t,\ell}^{[k]}, \end{aligned} \quad (11)$$

where $\mathbf{e}_1$ is the unit vector (the first column of $\mathbf{I}_D$), and where $(\mathbf{A}_t^{[k]})^{\mathsf{H}}\boldsymbol{\epsilon}^{[k]} = [\epsilon^{[k]};\mathbf{q}^{[k]}]$ is the transformed error vector, whose asymptotic order is also $o(1)$. The ISR of the kth extracted SOI (evaluated over all blocks and sub-blocks) is given by

$$\mathtt{ISR}^{[k]} = \frac{\left\langle \mathrm{E}\left[\left|(\mathbf{q}^{[k]})^{\mathsf{H}}\mathbf{z}_{t,\ell}^{[k]}\right|^2\right]\right\rangle_{t,\ell}}{|1+\epsilon^{[k]}|^2\left\langle \mathrm{E}\left[\left|s_{t,\ell}^{[k]}\right|^2\right]\right\rangle_{t,\ell}} = \frac{\mathtt{tr}\left[\left\langle \mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\right\rangle_{t,\ell}\mathbf{q}^{[k]}(\mathbf{q}^{[k]})^{\mathsf{H}}\right]}{\left\langle(\sigma_{t,\ell}^{[k]})^2\right\rangle_{t,\ell}} + o(1), \quad (12)$$

where $\mathbf{C}_{\mathbf{z},t,\ell}^{[k]} = \mathrm{E}[\mathbf{z}_{t,\ell}^{[k]}(\mathbf{z}_{t,\ell}^{[k]})^{\mathsf{H}}]$, $\mathtt{tr}[\cdot]$ denotes the trace of an argument, and $\langle\cdot\rangle_{t,\ell} = \langle\langle\cdot\rangle_{\ell}\rangle_t$.

The expression in (12) shows that the asymptotic value of ISR depends purely on $\mathbf{q}^{[k]}$, which can be considered as the lower part of $\widehat{\mathbf{w}}^{[k]}$ when the observed signals are actually unmixed, i.e., when $\mathbf{x}_{t,\ell}^{[k]} = \mathbf{u}_{t,\ell}^{[k]}$. This brings substantial simplifications of the following analyses because we can analyze the value of ISR when the true separating and mixing vectors are all unit vectors, that is, when $\mathbf{w}^{[k]} = \mathbf{a}_t^{[k]} = \mathbf{e}_1$; this property of ISR has been referred to as *equivariance*[7] [49]. Finally, the approximate mean value of ISR is given by

$$\mathrm{E}\left[\mathtt{ISR}^{[k]}\right] \approx \frac{\mathtt{tr}\left[\mathbf{C}_{\mathbf{z}}^{[k]}\mathtt{cov}[\mathbf{q}^{[k]}]\right]}{\sigma_k^2}, \quad (13)$$

where $\mathtt{cov}[\mathbf{q}^{[k]}] = \mathrm{E}[\mathbf{q}^{[k]}(\mathbf{q}^{[k]})^{\mathsf{H}}]$, $\mathbf{C}_{\mathbf{z}}^{[k]} = \left\langle\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\right\rangle_{t,\ell}$ and $\sigma_k^2 = \left\langle(\sigma_{t,\ell}^{[k]})^2\right\rangle_{t,\ell}$.

VI. Cramér-Rao Lower Bound

The lower bound for mean ISR shall now be obtained when $\mathtt{cov}[\mathbf{q}^{[k]}]$ is replaced by the CRLB for the corresponding parameter, which is the lower part of the mixing vector $\mathbf{h}^{[k]}$. We now remind the general approach to derive the CRLB for complex-valued parameters, which is then applied to our model. In the derivations, we can follow some results of Section IV.E in [11], knowing that the current model is an extension for $K > 1$ and $L > 1$.

According to [50], for any unbiased estimator of a parameter vector $\boldsymbol{\theta}$, it holds that

$$\mathtt{cov}\left[\boldsymbol{\theta}\right] \succeq \mathcal{J}^{-1}(\boldsymbol{\theta}) = \mathrm{CRLB}(\boldsymbol{\theta}), \quad (14)$$

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

$$\mathcal{J}(\boldsymbol{\theta}) = \begin{pmatrix}\mathbf{F} & \mathbf{P}\\ \mathbf{P}^* & \mathbf{F}^*\end{pmatrix}, \quad (15)$$

where

$$\mathbf{F} = \mathrm{E}\left[\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}^*}\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}^*}\right)^{\mathsf{H}}\right], \quad \mathbf{P} = \mathrm{E}\left[\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}^*}\left(\frac{\partial\mathcal{L}}{\partial\boldsymbol{\theta}^*}\right)^{\mathsf{T}}\right], \quad (16)$$

[6] In practice, the SOI uncertainty must be resolved through additional information to make the results of the analyses applicable. However, this is a known problem that is beyond the scope of this paper; see, e.g., [47], [48].

[7] More precisely, by definition, the vector $\mathbf{q}^{[k]}$ is time-varying (dependent on t). However, if, due to equivariance, we consider $\mathbf{w}^{[k]} = \mathbf{a}_t^{[k]} = \mathbf{e}_1$, it is time-invariant; we omit this detail in order to simplify the exposition.

and where the derivatives in (16) are defined according to the Wirtinger calculus; $\mathcal{L}$ denotes the log-likelihood function for the given estimation problem.

The model studied in this paper is parameterized by mixing and separating vectors. For computing the CRLB on the variance of their estimation, we have to cope with the uncertainties discussed in Section Section II-B. The SOI uncertainty does not appear in this analysis as there is the assumption of the true parameter values, which solves the problem itself. On the other hand, the scaling ambiguity needs to be treated so that the true parameter values are unambiguous. Since the target criterion, ISR is invariant to the scale of the output signal, we can simply assume $\beta^{[k]} = 1$.

Now, owing to the distortionless constraint $(\mathbf{w}^{[k]})^{\mathsf{H}}\mathbf{a}_t^{[k]} = 1$, it holds that $\gamma_1^{[K]}, \ldots, \gamma_T^{[k]}$ are dependent variables through $\gamma_t^{[k]} = 1 - (\mathbf{h}^{[k]})^{\mathsf{H}}\mathbf{g}_t^{[k]}$. Hence, the free parameters of the model are $\mathbf{g} = [\mathbf{g}^{[1]}; \ldots; \mathbf{g}^{[K]}]$ where $\mathbf{g}^{[k]} = [\mathbf{g}_1^{[k]}; \ldots; \mathbf{g}_T^{[k]}]$ and $\mathbf{h} = [\mathbf{h}^{[1]}; \ldots; \mathbf{h}^{[K]}]$. For computing the CRLB with (14), we therefore define the $K(T+1)(D-1)\times 1$ parameter vector $\boldsymbol{\theta}$ as

$$\boldsymbol{\theta} = \left[\mathbf{g}^{[1]}; \mathbf{h}^{[1]}; \ldots; \mathbf{g}^{[K]}; \mathbf{h}^{[K]}\right]. \tag{17}$$

For applying (13), we need to compute the diagonal block of $\mathrm{CRLB}(\boldsymbol{\theta})$ corresponding to $\mathbf{h}^{[k]}$. The detailed derivations are given in Appendix A, where we show that if the background signals $\mathbf{z}_{t,\ell}^{[k]}$ have zero mean circular Gaussian distribution with covariance $\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}$, which is the simplifying assumption introduced in Section IV-B, the CRLB is given by

$$\begin{aligned}&\mathtt{CRLB}(\mathbf{h}^{[k]})|_{\mathbf{h}=\mathbf{g}=0}\\ &= \frac{1}{N}\left\langle\left\langle \kappa_{s,t,\ell}^{[k]}\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\right\rangle_\ell - \left\langle (\sigma_{t,\ell}^{[k]})^2 (\mathbf{C}_{\mathbf{z},t,\ell}^{[k]})^{-1}\right\rangle_\ell^{-1}\right\rangle_t^{-1}.\end{aligned} \tag{18}$$

By replacing $\mathtt{cov}[\mathbf{q}^{[k]}]$ in (13), we obtain so-called Cramér-Rao-induced lower bound (CRIB) for ISR, which says that

$$\begin{aligned}\mathrm{E}\left[\mathtt{ISR}^{[k]}\right] \geq \frac{1}{N}\mathtt{tr}\Bigg(&\frac{\mathbf{C}_{\mathbf{z}}^{[k]}}{\sigma_k^2}\times\\ &\left\langle\left\langle \kappa_{s,t,\ell}^{[k]}\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\right\rangle_\ell - \left\langle (\sigma_{t,\ell}^{[k]})^2 (\mathbf{C}_{\mathbf{z},t,\ell}^{[k]})^{-1}\right\rangle_\ell^{-1}\right\rangle_t^{-1}\Bigg),\end{aligned} \tag{19}$$

where

$$\kappa_{s,t,\ell}^{[k]} = \mathrm{E}\left[|\psi_{t,\ell}^{[k]}(\mathbf{s}_{t,\ell})|^2\right] \tag{20}$$

and

$$\psi_{t,\ell}^{[k]}(\mathbf{s}_{t,\ell}) = -\frac{\partial \log p_{s_{t,\ell}}(\mathbf{s}_{t,\ell})}{\partial s_k} \tag{21}$$

is the kth score function corresponding to the pdf of the SOI. This result extends the bound given by [11, Eq. (70)] for joint extraction ($K > 1$) and for nonstationary (also Gaussian) SOI ($L > 1$). Detailed discussions are provided in Section VIII.

VII. Performance Analysis

This section formulates the main result of the analysis of mean ISR obtained by the solution corresponding to the optimum point of the contrast function (10). It should correspond to the average ISR achieved by algorithms optimizing this contrast, provided that they converge to the correct stationary point each time. However, the algorithms in [2] optimize (10) under the orthogonal constraint

$$\mathbf{a}_t^{[k]} = \frac{\widehat{\mathbf{C}}_t^{[k]}\mathbf{w}^{[k]}}{(\mathbf{w}^{[k]})^{\mathsf{H}}\widehat{\mathbf{C}}_t^{[k]}\mathbf{w}^{[k]}}, \tag{22}$$

where $\widehat{\mathbf{C}}_t^{[k]} = \langle\widehat{\mathbf{C}}_{t,\ell}^{[k]}\rangle_\ell$ and $\widehat{\mathbf{C}}_{t,\ell}^{[k]} = \hat{\mathrm{E}}[\mathbf{x}_{t,\ell}^{[k]}(\mathbf{x}_{t,\ell}^{[k]})^{\mathsf{H}}]$; see [2, Section III.A]. The constraint is introduced to ensure stable convergence and must be taken into account in the analysis. The following lemma formulates the main result.

Lemma 1: Let $\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}$ and $(\sigma_{t,\ell}^{[k]})^2$, respectively, be the covariance matrix of the background signals $\mathbf{z}^{[k]}$ and the variance of the SOI on the tth block, the lth sub-block, and the kth mixture. Let

$$\varphi_{t,\ell}^{[k]} = \mathrm{E}\left[\left|\phi^{[k]}\left(\left\{\frac{s_{t,\ell}^{[k]}}{\sigma_{t,\ell}^{[k]}}\right\}_k\right)\right|^2\right], \tag{23}$$

$$\nu_{t,\ell}^{[k]} = \mathrm{E}\left[\phi^{[k]}\left(\left\{\frac{s_{t,\ell}^{[k]}}{\sigma_{t,\ell}^{[k]}}\right\}_k\right)\frac{s_{t,\ell}^{[k]}}{\sigma_{t,\ell}^{[k]}}\right], \tag{24}$$

$$\rho_{t,\ell}^{[k]} = \mathrm{E}\left[\frac{\partial\phi^{[k]}}{\partial s_k^*}\left(\left\{\frac{s_{t,\ell}^{[k]}}{\sigma_{t,\ell}^{[k]}}\right\}_k\right)\right], \tag{25}$$

where $\phi^{[k]}$ (we neglect the subscripts t and ℓ for simplicity) is the kth score function related to the model density $f(\cdot)$ in (10) defined as

$$\phi^{[k]}\left(\left\{\frac{s_{t,\ell}^{[k]}}{\sigma_{t,\ell}^{[k]}}\right\}_k\right) = -\frac{\partial}{\partial s_k}\log f\left(\left\{\frac{s_{t,\ell}^{[k]}}{\sigma_{t,\ell}^{[k]}}\right\}_k\right), \tag{26}$$

where the partial derivative is taken over the kth argument denoted by s_k. Then, in the kth mixture, the mean ISR corresponding to the optimum point of the contrast function (10) in the vicinity of the true $\mathbf{w}^{[k]}$ under the orthogonal constraint (22) is given by (13), where

$$\mathtt{cov}[\mathbf{q}^{[k]}] = \frac{1}{N}\mathbf{R}^{[k]}\mathbf{S}^{[k]}(\mathbf{R}^{[k]})^{\mathsf{H}} + o(N^{-1}), \tag{27}$$

and

$$\begin{aligned}\mathbf{S}^{[k]} = \Bigg\langle &\frac{\langle(\sigma_{t,\ell}^{[k]})^2\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\rangle_\ell}{\langle(\sigma_{t,\ell}^{[k]})^2\rangle_\ell^2} + \left\langle\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\frac{\varphi_{t,\ell}^{[k]}}{(\sigma_{t,\ell}^{[k]})^2|\nu_{t,\ell}^{[k]}|^2}\right\rangle_\ell\\ &- 2\frac{\langle\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\rangle_\ell}{\langle(\sigma_{t,\ell}^{[k]})^2\rangle_\ell}\Bigg\rangle_t,\end{aligned} \tag{28}$$

$$\mathbf{R}^{[k]} = \left\langle\frac{\langle\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\rangle_\ell}{\langle(\sigma_{t,\ell}^{[k]})^2\rangle_\ell} - \left\langle\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\frac{\rho_{t,\ell}^{[k]}}{(\sigma_{t,\ell}^{[k]})^2\nu_{t,\ell}^{[k]}}\right\rangle_\ell\right\rangle_t^{-1}. \tag{29}$$

Proof: See Appendix B. ∎

Before discussing some special cases, it is worth noting that the result (27) agrees with [9, Eq. (89)] when $L = 1$, which correctly follows the previous model from [9].

TABLE III
OVERVIEW OF THEORETICAL BOUNDS

model	setting	CRLB on ISR	perf. anal.	efc.
ng. ICE	$L=T=K=1$	[24]	[4], [19]	✓
nst. ICE	$T=K=1$, $L>1$	here	here	✗
ng. IVE	$L=T=1$, $K>1$	[27]	here	✓
nst. IVE	$T=1$, $L>1$, $K>1$	here	here	✗
ng. CSV	$L=1$, $T>1$, $K>1$	[11]	[9]	✓
nst. CSV	$L>1$, $T>1$, $K>1$	here	here	✗

VIII. DISCUSSION

A. Previous bounds

The proposed model is an extension of the previous BSE[8] models ICE, IVE, and CSV. These special cases have already been discussed in previous literature [9], [11], so here we limit ourselves to a brief summary in Table III. The model acronyms are consistent with those introduced in the Introduction, and "nG." and "nst." correspond to the non-Gaussianity and non-stationarity source models. The references in the table correspond to analyses similar to those in this paper, although some are restricted to the real-valued case, while our analyses include the complex-valued case, non-circular SOI, and $K>1$. The table thus illustrates how comprehensive the double non-stationarity model is and hence the impact of the analyses presented in this paper.

An important situation is when the model pdf (9) matches the true (normalized) pdf, in which case $\nu_{t,\ell}^{[k]}=1$ and $\varphi_{t,\ell}^{[k]}=\rho_{t,\ell}^{[k]}=\overline{\kappa}_{s,t,\ell}^{[k]}$, where $\overline{\kappa}_{s,t,\ell}^{[k]}=\kappa_{s,t,\ell}^{[k]}(\sigma_{t,\ell}^{[k]})^2$ corresponds to $\kappa_{s,t,\ell}^{[k]}$ of the normalized (unit variance) pdf. In that case, the contrast function approaches the likelihood function, and the estimates obtained based on its optimization can achieve statistical efficiency (the "efc." acronym in Table III). This occurs if the leading term in (27) and (18) coincide.

The reader can easily verify that (27) and (18) do *not* generally coincide in this special case, as happens for the non-Gaussianity-based source model, i.e., when $L=1$ (see [9]). The bounds coincide with stronger assumption when $\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}$ is independent of ℓ. Consequently, the algorithms from [2] are generally not asymptotically efficient unless the background signals are stationary within blocks. The problem of how to achieve statistical optimality also for non-stationary background is beyond the scope of this paper and can be a suitable topic for further research[9].

B. Gaussian SOI: Identifiability

In this section, we focus our discussion on the case when the marginal pdf of the SOI, i.e. for all k, is Gaussian, in order to reveal how non-stationarity and non-circularity contribute to the identifiability of the SOI. For the Gaussian case, it can be shown that

$$\kappa_{s,t,\ell}^{[k]}=\Big((\sigma_{t,\ell}^{[k]})^2\big(1-|\overline{\lambda}_{t,\ell}^{[k]}|^2\big)\Big)^{-1}, \tag{30}$$

where $\overline{\lambda}_{t,\ell}^{[k]}=(\sigma_{t,\ell}^{[k]})^{-2}\mathrm{E}\big[(s_{t,\ell}^{[k]})^2\big]$ corresponds to a normalized non-circularity coefficient of $s_{t,\ell}^{[k]}$, whose values range from $[0,1]$; $s_{t,\ell}^{[k]}$ is circular when $\overline{\lambda}_{t,\ell}^{[k]}=0$. Next, we assume that the background is stationary over all blocks and sub-blocks, which means that $\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}=\mathbf{C}_{\mathbf{z}}^{[k]}$. This together with (30) yields that the CRLB (19) turns to

$$\mathrm{E}\left[\mathtt{ISR}^{[k]}\right]\geq\frac{1}{N}\frac{d-1}{\sigma_k^2}\times \left(\left\langle\frac{1}{(\sigma_{t,\ell}^{[k]})^2\big(1-|\overline{\lambda}_{t,\ell}^{[k]}|^2\big)}\right\rangle_{t,\ell}-\left\langle\frac{1}{\langle(\sigma_{t,\ell}^{[k]})^2\rangle_\ell}\right\rangle_t\right)^{-1} \tag{31}$$

The SOI is not identifiable when the right-hand side of (31) is infinite, which happens when the expression in the parenthesis is zero. To simplify, let us assume for now that $K=T=1$ when we can omit the indices k and t, and (19) simplifies to

$$\mathrm{E}\left[\mathtt{ISR}^{[k]}\right]\geq\frac{1}{N}\frac{d-1}{\sigma_k^2}\left(\left\langle\frac{1}{\sigma_\ell^2\big(1-|\overline{\lambda}_\ell|^2\big)}\right\rangle_\ell-\frac{1}{\langle\sigma_\ell^2\rangle_\ell}\right)^{-1}. \tag{32}$$

In this expression, it is easier to see that the SOI is always identifiable when $\overline{\lambda}_\ell\neq 0$ for any ℓ, so the non-circularity of the SOI even in a single sub-block is sufficient for the SOI to be identifiable. For the circular SOI in all sub-blocks, we see that the expression in the brackets compares the harmonic mean with the average value of the SOI's variance. The averages are known to be equal if and only if the variance is constant (the SOI is stationary). This shows that the non-stationarity also strongly contributes to the SOI's identifiability.

One more case worth considering is when the SOI is Circular Gaussian and stationary (i.e. over blocks as well as sub-block), while the background is non-stationary. Then, (19) has the form

$$\mathrm{E}\left[\mathtt{ISR}^{[k]}\right]\geq\frac{1}{N}\mathtt{tr}\left(\mathbf{C}_{\mathbf{z}}^{[k]}\left\langle\left\langle\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\right\rangle_\ell-\left\langle\big(\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}\big)^{-1}\right\rangle_\ell^{-1}\right\rangle_t^{-1}\right). \tag{33}$$

This expression shows that, in this case, for the identifiability of the SOI, the background must be non-stationary within any sub-block. On the other hand, pure non-stationarity of the background over blocks (while stationarity within sub-block) does not guarantee identifiability.

C. Multidimensional Circular Gaussian SOI: $K=2$

In the joint extraction setting, i.e. when $K>1$, the dependencies within the vector component of the SOI play an important role. They are expected to help with solving the permutation problem as well as improving the extraction accuracy. The analyses of the double nonstationarity model enable us to fully investigate the $K=2$ real-valued case when the joint pdf of the SOI is Gaussian. For simplicity, we consider $T=L=1$, in which case identifiability and extraction accuracy depend purely on the correlation coefficient between the two SOI components $s^{[1]}$ and $s^{[2]}$.

Let $\mathbf{s}\sim\mathcal{N}(\mathbf{0},\mathbf{\Sigma}_\xi)$ where

$$\mathbf{\Sigma}_\xi=\begin{pmatrix}1&\xi\\ \xi&1\end{pmatrix} \tag{34}$$

[8]Let us remind the reader of a brief overview of the continuity with previous models in the third paragraph of Section I.

[9]After a detailed inspection of the proof in Appendix B we conjecture that the orthogonal constraint (22) causes this performance suboptimality as it does not reflect varying background covariance over sub-blocks.

is the covariance of the SOI, and $\xi \in [0,1)$ is the true correlation coefficient between the SOI's components. Let $\widehat{\xi} \in [0,1)$ be the correlation coefficient that we select in the model pdf, which is also Gaussian. By the definitions (20) and (23)-(25), it is easy to show that, in this special case,

$$\kappa^{[k]} = \frac{1}{1-\xi^2}, \qquad \varphi^{[k]} = \frac{1-2\xi\widehat{\xi}+\widehat{\xi}^2}{(1-\widehat{\xi}^2)^2}, \tag{35}$$

$$\nu^{[k]} = \frac{1-\xi\widehat{\xi}}{1-\widehat{\xi}^2}, \qquad \rho^{[k]} = \frac{1}{1-\widehat{\xi}^2}, \tag{36}$$

for $k \in \{1,2\}$. Now, by putting into (19) and (27), which are now both independent of $\mathbf{C}_{\mathbf{z}}^{[k]}$, it is easy to compute that the CRIB as well as the performance analysis coincide and say that

$$\mathrm{E}\left[\mathtt{ISR}^{[k]}\right] \geq \frac{d-1}{N} \cdot \frac{1-\xi^2}{\xi^2}. \tag{37}$$

This result makes a clear sense in that the mean ISR grows to infinity as ξ approaches zero, i.e., when the correlation between mixtures is vanishing. What is surprising is the fact that (27) does not depend on $\widehat{\xi}$! This tells us that it does not matter at all how well we model the correlation between the SOI components (only the values $\widehat{\xi}=0$ and $\widehat{\xi}=1$ are not allowed).

We experimentally examine a similar scenario for $K>2$ in Section IX-B.

IX. Experimental Validation

This section is devoted to experimental validation of the derived analyses. The first example verifies the results when $L>1$, i.e. in the extension provided by the double non-stationarity model (the validity of the analyses for the case $L=1$ follows from the previous literature due to the analytical coincidence of the results). In the second experiment, we focus on the case when $K>1$, where the analyses provide new insight into the impact of correctly modeling the dependencies between mixtures. The third experiment focuses on the possible use of performance analysis to post-estimate the achieved ISR, the value of which cannot be determined under realistic blind extraction conditions. The Matlab implementations are available on-line[10].

A. Validation in a Double Nonstationarity Setting

To perform a validation of (19) and (27), we repeat the numerical simulation done in [9, Section V.A-2]. The parameters of a mixture are $d=6$, $L=5$, $T=3$, and $K=1$. The length of data N varies from 150 through 150000, which means that N_s ranges from the extremely small value of $N_s=15$ through $N_s=1,000$. The SOI is non-stationary and non-circular Gaussian with non-stationarity parameter $\alpha=2$ (see [9, Section V.A-2]) and $\overline{\lambda}=0.5$. The background signals are non-stationary circular Gaussian over blocks while stationary over sub-blocks.

Figure 3 shows averaged ISRs achieved after $1,000$ trials by the selected algorithms, which is compared to their theoretical counterparts (denoted by "teo") evaluated through (19) and (27). "FastDIVA rati $L=1$" corresponds to the original algorithm from [9] where each block of the SOI is modeled as non-Gaussian and stationary; the nonlinearity $\phi(s)=\frac{s^*}{1+|s|^2}$ denoted as "rati" is used. Its theoretical performance is estimated using (27) where the statistics are computed using sample averages over the samples within the blocks. Despite the discrepancy between the assumed ($L=1$) and true ($L=5$) source models, the theoretical and empirical ISRs are in good agreement. "FastDIVA rati $L=5$" corresponds to the extension exploiting the double non-stationarity model with $L=5$. The results show the benefit of the proper modeling of the SOI's non-stationarity, however, the improvement is only by about 2 dB, because both algorithms do not take non-circularity into account. "FastDIVA gauss $L=5$" assumes the correct Gaussian model of the SOI, including non-circularity. Its performance is, therefore, significantly better and does not drop so much for small values of N_s. Moreover, the method attains the CRiB, which, in the example here, coincides with the ISR predicted by the performance analysis because the background is stationary over sub-blocks.

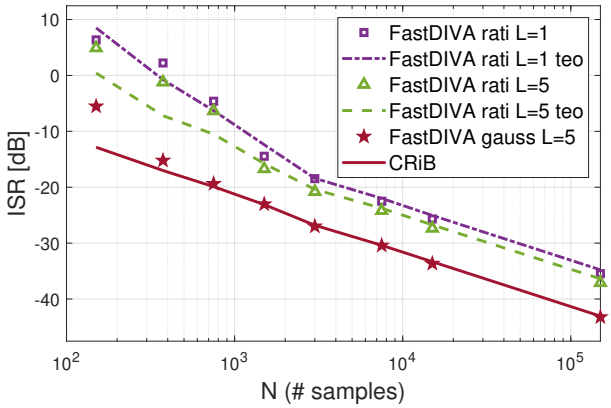

Fig. 3. ISR averaged over 1000 trials as a function of the data length N. The theoretical ISR of the algorithms is computed using (27); the CRiB is computed according to (19). The 5% trimmed mean is used for averaging the ISRs achieved by the algorithms in order to avoid trials where the given algorithm extracts a different independent source than the SOI. Owing to the ambiguity of the BSE task, these cases do not necessarily mean failures.

B. Modelling the Dependencies by Adjacent Correlations

The experiment here studies a similar problem to that of Section VIII-C. Joint extraction of $K=5$ real-valued mixtures with $d=10$, $T=1$, $L=5$, and $N=5000$ is considered. The SOI is stationary Gaussian having the tri-diagonal covariance matrix

$$\mathbf{\Sigma} = \begin{pmatrix} 1 & \xi & & \\ \xi & 1 & \ddots & \\ & \ddots & \ddots & \xi \\ & & \xi & 1 \end{pmatrix}, \tag{38}$$

where ξ represents the correlation coefficient of adjacent SOI components. The background signals are Gaussian and stationary. Since all signals in this scenario are stationary, the model is identifiable only for $\xi>0$.

[10] https://github.com/koldovsk/Double-Nonstationarity-Model-Analysis

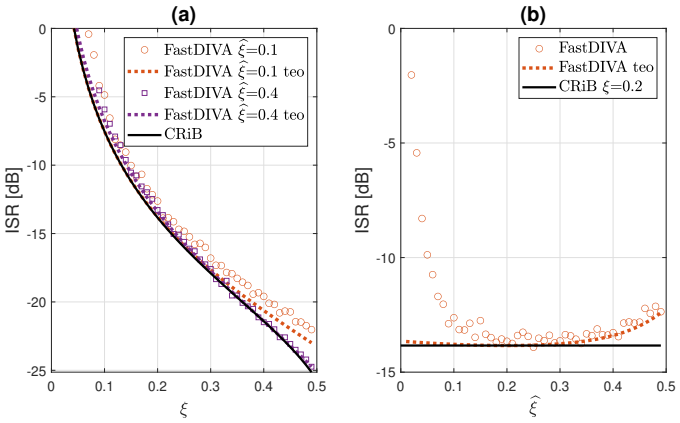

Fig. 4. Average ISR and the CRiB after 100 trials when (a) the true correlation coefficient varies $\xi \in [0, 0.5]$, and when (b) $\xi = 0.2$ and the model correlation coefficient assumed by FastDIVA varies over $\widehat{\xi} \in [0, 0.5]$.

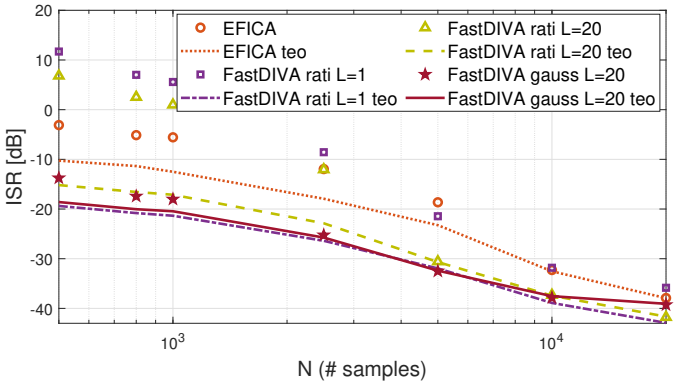

Fig. 5. 5%-trimmed mean ISR evaluated over 1000 trials as a function of the data length N. The SOI is a random piece of utterance. The empirical ISR of the methods is compared with its theoretical post-estimate (denoted as "teo").

The CRiB and the performance of FastDIVA are compared for $\xi \in [0, 0.5]$, in which range the covariance matrix $\mathbf{\Sigma}$ is positive definite and well conditioned. FastDIVA operates with the tri-diagonal covariance matrix model but with a correlation coefficient equal to $\widehat{\xi} \in [0, 0.5]$. The results are shown in Figure 4(a) for varying ξ and Figure 4(b) for varying $\widehat{\xi}$.

The results confirm the validity of (19) and (27) in the sense that the CRiB is always lower or equal to (27) and they coincide when $\xi = \widehat{\xi}$. The CRiB grows to infinity as ξ approaches zero, which agrees with the identifiability condition of this scenario. The empirical ISR by FastDIVA is in good agreement with (27) up to the cases when $\widehat{\xi}$ is close to zero (Figure 4(b)). The discrete points at which the empirical ISR reached or exceeded the CRiB are due to statistical errors.

There is an interesting phenomenon worthy of mention: The loss of performance of FastDIVA due to a mismatched correlation coefficient ($\widehat{\xi} \not\approx \xi$) is mostly negligible unless $\widehat{\xi}$ is too close to zero (an underestimated ξ). Such a result indicates certain robustness of (j)BSE to inaccuracies in the source model.

C. Interference-to-Signal Ratio Post-Estimation

In the blind scenario, the achieved ISR cannot be measured. However, such information would be valuable, allowing us to make the right decisions on how to deal with the result of BSE. The theoretical ISR given by relation (27) has interesting potential in this respect: It can be used to post-estimate the achieved (mean) ISR when replacing the unknown statistics with their counterparts estimated from the extracted signal.

This approach has proved useful in selecting the best-fitting source model in [51]. However, significant bias is observed for the non-Gaussian model when processing real-world signals that do not obey this model exactly, particularly when they are non-stationary.

In the experiment here, we use a small data set of 15 recordings of Czech utterances, each of length 8.5 s sampled at 16 kHz. In a trial, a random piece of utterance of length N is randomly mixed with a stationary Gaussian background to generate an instantaneous mixture of dimension $d = 6$; we consider here the standard (real-valued) ICE setup, i.e., with $T = K = 1$. The speech signal plays the role of the SOI. Four methods are compared by which the SOI can be extracted: EFICA [31] and FastDIVA-rati-1, which are purely non-Gaussianity-based; FastDIVA-rati-20, which also exploits non-stationarity; FastDIVA-gauss-20, which is purely non-stationarity-based. EFICA separates the mixture into d independent components (it performs ICA); the one corresponding to the SOI is selected as the output.

The ISR achieved by the methods is post-estimated by using corresponding formulas of the theoretical mean ISR and compared with the empirical one. Averaged values after 1000 trials are shown in Figure 5 as functions of the data length N; the averages were computed using the 5%-trimmed mean to disregard outliers caused by the big variability of the SOI.

For sufficiently large data ($N \geq 5000$), we observe a good correspondence between the empirical and post-estimated ISR. However, for $N < 5000$, the theoretical ISR of the methods using non-Gaussianity is already significantly biased (overly optimistic). FastDIVA using a Gaussian model with $L = 20$ not only achieves the best performance but also its theoretical ISR agrees well with the empirical one.

X. Conclusions

The theoretical mean ISR achievable through the double non-stationarity model has been analyzed through the CRLB and through the performance analysis. The experiments validate these results showing the superiority of the double nonstationarity model over the previous models. The analyses reveal identifiability conditions and the achievable and practically attained extraction accuracy, showing which signal properties influence these aspects and how. In particular, the advantages of the Gaussian-only model have been analyzed. Lastly, the ISR post-estimation has been validated, particularly for the Gaussian-only model, which achieves little bias and can be very useful for decision-making.

Acknowledgements

We would like to thank Dr. Dana Lahat for her significant help with this article. She spent a lot of time on discussions and a very careful review of the entire article.

APPENDIX A
PROOF OF (18)

The next step is to compute the FIM according to (16) where $\mathcal{L}$ is given by (8). The FIM is a square matrix of dimension $2K(T+1)(D-1)$. Owing to the structure of $\boldsymbol{\theta}$, we partition the FIM into $2K(T+1)\times 2K(T+1)$ blocks each of dimension $(D-1)\times(D-1)$ corresponding either to $\mathbf{g}_t^{[k]}$ or $\mathbf{h}^{[k]}$ or to their conjugate counterparts. We also exploit the additive structure of (8) and compute first the Wirtinger derivatives of (7), and evaluate them for the special case that $\mathbf{w}^{[k]}=\mathbf{a}_t^{[k]}=\mathbf{e}_1$.

By comparing (7) with [11, Eq. 56], we can see that the log-likelihood functions are algebraically similar up to the fact that (7) involves the determinant term equal to $\sum_{k=1}^{K}(D-2)\log|1-(\mathbf{h}^{[k]})^{\mathsf{H}}\mathbf{g}_t^{[k]}|^2$. By this, the Wirtinger derivatives of (7) by $\mathbf{g}_t^{[k]}$ and by $\mathbf{h}^{[k]}$ easily follow from [11, Eq. 57] and [11, Eq. 58], respectively, provided that the derivatives of the determinant term are equal to zero for $\mathbf{g}=\mathbf{h}=\mathbf{0}$. We, therefore, focus on proving the latter.

To this end, we compute the following derivatives:

$$\frac{\partial}{\partial(\mathbf{g}_t^{[k]})^*}\log|1-(\mathbf{h}^{[k]})^{\mathsf{H}}\mathbf{g}_t^{[k]}|^2=\frac{-\mathbf{h}^{[k]}}{|1-(\mathbf{h}^{[k]})^{\mathsf{H}}\mathbf{g}_t^{[k]}|^2}, \tag{41}$$

$$\frac{\partial}{\partial(\mathbf{h}^{[k]})^*}\log|1-(\mathbf{h}^{[k]})^{\mathsf{H}}\mathbf{g}_t^{[k]}|^2=\frac{-\mathbf{g}_t^{[k]}}{|1-(\mathbf{h}^{[k]})^{\mathsf{H}}\mathbf{g}_t^{[k]}|^2}. \tag{42}$$

Considering the special case that $\mathbf{g}=\mathbf{h}=\mathbf{0}$, we can see that these terms are equal to zero. Hence, it holds that

$$\frac{\partial\mathcal{L}_{t,\ell}(\mathbf{a}_t,\mathbf{w}|\mathbf{x}_{t,\ell})}{\partial(\mathbf{g}_{\tilde{t}}^{[\tilde{k}]})^*}\bigg|_{\mathbf{g}=\mathbf{h}=\mathbf{0}}=-\delta_{k\tilde{k}}\,\delta_{t\tilde{t}}\,(s_{t,\ell}^{[k]})^*(\boldsymbol{\pi}_{t,\ell}^{[k]})^*(\mathbf{z}_{t,\ell}^{[k]}), \tag{43a}$$

$$\frac{\partial\mathcal{L}_{t,\ell}(\mathbf{a}_t,\mathbf{w}|\mathbf{x}_{t,\ell})}{\partial(\mathbf{h}^{[\tilde{k}]})^*}\bigg|_{\mathbf{g}=\mathbf{h}=\mathbf{0}}=\delta_{k\tilde{k}}\,\psi_{t,\ell}^{[k]}(s_{t,\ell})\mathbf{z}_{t,\ell}^{[k]}, \tag{43b}$$

where $\psi_{t,\ell}^{[k]}$ is given by (21), $\boldsymbol{\pi}_{t,\ell}^{[k]}=-\frac{\partial\log p_{\mathbf{z}_{t,\ell}^{[k]}}}{\partial\mathbf{z}}$, $\delta_{k\tilde{k}}$ stands for the Kronecker delta, $\tilde{k}=1,\ldots,K$, and $\tilde{t}=1,\ldots,T$; (43a) determines the pqth block of $\mathbf{F}$ where $p=(k-1)(T+1)+t$ and $q=(\tilde{k}-1)(T+1)+\tilde{t}$, and (43b) determines the block for $p=k(T+1)$ and $q=\tilde{k}(T+1)$. The Kronecker deltas in (43a) and (43b) cause that $\mathbf{F}$ and $\mathbf{P}$ are block-diagonal as we show in the following.

Let us now introduce a $(T+1)(D-1)\times 1$ vector of derivatives:

$$\boldsymbol{\nabla}_{t,\ell}^{[k]}=\Big[\underbrace{\mathbf{0};\ldots;\mathbf{0}}_{t-1};-(s_{t,\ell}^{[k]})^*(\boldsymbol{\pi}_{t,\ell}^{[k]})^*;\underbrace{\mathbf{0};\ldots;\mathbf{0}}_{T-t};\psi_{t,\ell}^{[k]}\mathbf{z}_{t,\ell}^{[k]}\Big], \tag{44}$$

where we write the score functions without arguments to simplify the notation, and auxiliary matrices:

$$\mathbf{F}_{t,\ell}^{[k]}=\mathrm{E}\left[\boldsymbol{\nabla}_{t,\ell}^{[k]}(\boldsymbol{\nabla}_{t,\ell}^{[k]})^{\mathsf{H}}\right],\quad \mathbf{P}_{t,\ell}^{[k]}=\mathrm{E}\left[\boldsymbol{\nabla}_{t,\ell}^{[k]}(\boldsymbol{\nabla}_{t,\ell}^{[k]})^{\mathsf{T}}\right]. \tag{45}$$

By introducing

$$\mathbf{F}^{[k]}=N\cdot\langle\mathbf{F}_{t,\ell}^{[k]}\rangle_{t,\ell},\qquad \mathbf{P}^{[k]}=N\cdot\langle\mathbf{P}_{t,\ell}^{[k]}\rangle_{t,\ell}, \tag{46}$$

we can write the main blocks of $\mathcal{J}(\boldsymbol{\theta})$ as

$$\mathbf{F}=\mathtt{bdiag}\big(\mathbf{F}^{[1]},\ldots,\mathbf{F}^{[K]}\big),\quad \mathbf{P}=\mathtt{bdiag}\big(\mathbf{P}^{[1]},\ldots,\mathbf{P}^{[K]}\big), \tag{47}$$

where $\mathtt{bdiag}(\cdot)$ denotes a block-diagonal matrix having the argument on its main diagonal. By (45) and (46), the structures of $\mathbf{F}^{[k]}$ and $\mathbf{P}^{[k]}$ are, respectively, given by (39) and (40), where $\kappa_{s,t,\ell}^{[k]}$ is defined by (20) and

$$\boldsymbol{\kappa}_{\mathbf{z},t,\ell}^{[k]}=\mathrm{E}\left[\boldsymbol{\pi}_{t,\ell}^{[k]}\left(\boldsymbol{\pi}_{t,\ell}^{[k]}\right)^{\mathsf{H}}\right]^*. \tag{48}$$

Note that, in (40), we use $\mathbf{b}^2$ as a short notation of $\mathbf{b}\mathbf{b}^{\mathsf{T}}$ for a vector quantity $\mathbf{b}$. For a check, the reader can compare (39) and (40) with [11, Eq. (67)] and [11, Eq. (68)], respectively, as the studied model in [11] corresponds to the case $K=L=1$.

We now turn to the special case that the background signals $\mathbf{z}_{t,\ell}^{[k]}$ have zero mean circular Gaussian distribution with covariance $\mathbf{C}_{\mathbf{z},t,\ell}^{[k]}$. The corresponding score function then equals $\boldsymbol{\pi}_{t,\ell}^{[k]}(\mathbf{z})=\big((\mathbf{C}_{\mathbf{z},t,\ell}^{[k]})^{-1}\mathbf{z}\big)^*$, and, by (48), $\boldsymbol{\kappa}_{\mathbf{z},t,\ell}^{[k]}=(\mathbf{C}_{\mathbf{z},t,\ell}^{[k]})^{-1}$, and $\mathbf{P}^{[k]}=\mathbf{0}$ due to circularity. By (15), the whole FIM is block-diagonal, and $\mathtt{CRLB}(\mathbf{h}^{[k]})|_{\mathbf{h}=\mathbf{g}=0}$ is obtained as the lower right-corner $(D-1)\times(D-1)$ block of $N^{-1}(\mathbf{F}^{[k]})^{-1}$. By applying the block inverse-matrix lemma to (39), (18) follows, which concludes the proof. ∎

APPENDIX B
PROOF OF LEMMA 1

Let $\widehat{\mathbf{w}}^{[k]}$ denote the stationary point of (10) under the constraint (22), and let $\widehat{s}_{t,\ell}^{[k]}=(\widehat{\mathbf{w}}^{[k]})^{\mathsf{H}}\mathbf{x}_{t,\ell}^{[k]}$, $\widehat{\sigma}_{t,\ell}^{[k]}$ and $\widehat{\mathbf{z}}_t^{[k]}$ denote, respectively, the associated estimates of $s_{t,\ell}^{[k]}$, $\sigma_{t,\ell}^{[k]}$ and $\mathbf{z}_t^{[k]}$. We take advantage of the equivariance property of ISR and perform the analysis for the special case when $\mathbf{w}^{[k]}=\mathbf{a}_t^{[k]}=\mathbf{e}_1$. Therefore, the leading term in the asymptotic expression of $\widehat{\mathbf{w}}^{[k]}$ is sought in the form $\widehat{\mathbf{w}}^{[k]}\approx[1;\mathbf{q}^{[k]}]$. As the second step, $\mathbf{q}^{[k]}$ will be put into (13), which leads to the desired expression of the mean ISR.

Now, $\mathbf{q}^{[k]}$ will be computed as the asymptotic solution of $\nabla^{[k]}=\mathbf{0}$, where $\nabla^{[k]}$ is the gradient of (10) under (22) with respect to $\mathbf{w}^{[k]}$, which is, by [2, Eq. (38)] equal to

$$\nabla^{[k]}=\left\langle\widehat{\mathbf{a}}_t^{[k]}-\left\langle\frac{1}{\widehat{\nu}_{t,\ell}^{[k]}}\widehat{\mathrm{E}}\left[\phi^{[k]}\left(\left\{\frac{\widehat{s}_{t,\ell}^{[k]}}{\widehat{\sigma}_{t,\ell}^{[k]}}\right\}_k\right)\frac{\mathbf{x}_{t,\ell}^{[k]}}{\widehat{\sigma}_{t,\ell}^{[k]}}\right]\right\rangle_\ell\right\rangle_t, \tag{49}$$

where

$$\widehat{\nu}_{t,\ell}^{[k]}=\widehat{\mathrm{E}}\left[\phi^{[k]}\left(\left\{\frac{\widehat{s}_{t,\ell}^{[k]}}{\widehat{\sigma}_{t,\ell}^{[k]}}\right\}_k\right)\frac{\widehat{s}_{t,\ell}^{[k]}}{\widehat{\sigma}_{t,\ell}^{[k]}}\right], \tag{50}$$

and $\widehat{\mathbf{a}}_t^{[k]}$ is given by (22). Our goal is now to compute the asymptotic expansion of the right-hand side of (49); for further

$$\mathbf{F}^{[k]} = N \begin{pmatrix} \left\langle \boldsymbol{\kappa}^{[k]}_{\mathbf{z},1,\ell}(\sigma^{[k]}_{1,\ell})^2 \right\rangle_\ell & & \mathbf{0} & -\mathbf{I}_{D-1} \\ & \ddots & & \vdots \\ \mathbf{0} & & \left\langle \boldsymbol{\kappa}^{[k]}_{\mathbf{z},T,\ell}(\sigma^{[k]}_{T,\ell})^2 \right\rangle_\ell & -\mathbf{I}_{D-1} \\ -\mathbf{I}_{D-1} & \dots & -\mathbf{I}_{D-1} & \left\langle \kappa^{[k]}_{s,t,\ell}\mathbf{C}^{[k]}_{\mathbf{z},t,\ell} \right\rangle_{t,\ell} \end{pmatrix} \tag{39}$$

$$\mathbf{P}^{[k]} = N \cdot \mathtt{bdiag}\left(\left\langle \mathrm{E}\left[\left(\boldsymbol{\pi}^{[k]}_{1,\ell}\right)^2 \right]^* \mathrm{E}\left[\left(s^{[k]}_{1,\ell}\right)^2 \right]^* \right\rangle_\ell, \dots, \left\langle \mathrm{E}\left[\left(\boldsymbol{\pi}^{[k]}_{T,\ell}\right)^2 \right]^* \mathrm{E}\left[\left(s^{[k]}_{T,\ell}\right)^2 \right]^* \right\rangle_\ell, \left\langle \mathrm{E}\left[\left(\psi^{[k]}_{t,\ell}\right)^2 \right] \mathrm{E}\left[\left(\mathbf{z}^{[k]}_{t,\ell}\right)^2 \right] \right\rangle_{t,\ell} \right) \tag{40}$$

use, we introduce the following statistics

$$\widehat{\rho}^{[k]}_{t,\ell} = \widehat{\mathrm{E}}\left[\frac{\partial \phi^{[k]}}{\partial s^*_k} \left(\left\{ \frac{\widehat{s}^{[k]}_{t,\ell}}{\widehat{\sigma}^{[k]}_{t,\ell}} \right\}_k \right) \right], \tag{51}$$

$$\widehat{\boldsymbol{\chi}}^{[k]}_{t,\ell} = \widehat{\mathrm{E}}\left[\mathbf{z}^{[k]}_{t,\ell} (s^{[k]}_{t,\ell})^{\mathsf{H}} \right] \tag{52}$$

$$\widehat{\boldsymbol{\zeta}}^{[k]}_{t,\ell} = \widehat{\mathrm{E}}\left[\phi^{[k]} \left(\left\{ \frac{s^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right\}_k \right) \frac{\mathbf{z}^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right] \tag{53}$$

$$\widehat{\boldsymbol{\mu}}^{[k]}_{t,\ell} = \widehat{\mathrm{E}}\left[\phi^{[k]} \left(\left\{ \frac{\widehat{s}^{[k]}_{t,\ell}}{\widehat{\sigma}^{[k]}_{t,\ell}} \right\}_k \right) \frac{\mathbf{z}^{[k]}_{t,\ell}}{\widehat{\sigma}^{[k]}_{t,\ell}} \right]. \tag{54}$$

Now, since $\mathbf{w}^{[k]} = \mathbf{a}^{[k]}_t = \mathbf{e}_1$, by (2) and (4) it holds that $\mathbf{x}^{[k]}_{t,\ell} = [s^{[k]}_{t,\ell}; -\mathbf{z}^{[k]}_{t,\ell}]$, and we can write

$$\widehat{\mathbf{C}}^{[k]}_{t,\ell} = \widehat{\mathrm{E}}[\mathbf{x}^{[k]}_{t,\ell}(\mathbf{x}^{[k]}_{t,\ell})^{\mathsf{H}}] = \\ = \begin{bmatrix} (\sigma^{[k]}_{t,\ell})^2 + c^{[k]}_{t,\ell} & -(\widehat{\boldsymbol{\chi}}^{[k]}_{t,\ell})^{\mathsf{H}} \\ -\widehat{\boldsymbol{\chi}}^{[k]}_{t,\ell} & \mathbf{C}^{[k]}_{\mathbf{z},t,\ell} + \boldsymbol{\Xi}^{[k]}_{t,\ell} \end{bmatrix}, \tag{55}$$

where

$$c^{[k]}_{t,\ell} = \widehat{\mathrm{E}}[s^{[k]}_{t,\ell}(s^{[k]}_{t,\ell})^*] - (\sigma^{[k]}_{t,\ell})^2, \tag{56}$$

$$\boldsymbol{\Xi}^{[k]}_{t,\ell} = \widehat{\mathrm{E}}[\mathbf{z}^{[k]}_{t,\ell}(\mathbf{z}^{[k]}_{t,\ell})^{\mathsf{H}}] - \mathbf{C}^{[k]}_{\mathbf{z},t,\ell}. \tag{57}$$

It holds that

$$\frac{\widehat{s}^{[k]}_{t,\ell}}{\widehat{\sigma}^{[k]}_{t,\ell}} = \frac{(\widehat{\mathbf{w}}^{[k]})^{\mathsf{H}}\mathbf{x}^{[k]}_{t,\ell}}{\sqrt{(\sigma^{[k]}_{t,\ell})^2 + c^{[k]}_{t,\ell} - b^{[k]}_{t,\ell}}} = \\ = \frac{s^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} - \frac{(\mathbf{q}^{[k]})^{\mathsf{H}}\mathbf{z}^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} + o_p(N_s^{-1/2}), \tag{58}$$

where

$$b^{[k]}_{t,\ell} = \widehat{\mathrm{E}}[s^{[k]}_{t,\ell}(s^{[k]}_{t,\ell})^{\mathsf{H}}] - \widehat{\mathrm{E}}[\widehat{s}^{[k]}_{t,\ell}(\widehat{s}^{[k]}_{t,\ell})^{\mathsf{H}}] \tag{59}$$

Note that the stochastic order of $c^{[k]}_{t,\ell}$ and $\boldsymbol{\Xi}^{[k]}_{t,\ell}$ is the same as that of $\widehat{\boldsymbol{\chi}}^{[k]}_{t,\ell}$, i.e., $O_p(N_s^{-1/2})$, while $b^{[k]}_{t,\ell}$ is $O_p(N_s^{-1})$; $o_p(\cdot)$ and $O_p(\cdot)$ represent the stochastic-order symbols; see [52].

Next, by assuming that $\phi^{[k]}$ is sufficiently smooth, the first-order Taylor series expansion gives

$$\phi^{[k]}\left(\left\{ \frac{\widehat{s}^{[k]}_{t,\ell}}{\widehat{\sigma}^{[k]}_{t,\ell}} \right\}_k \right) = \phi^{[k]}\left(\left\{ \frac{s^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right\}_k \right) \\ - \sum_{k'=1}^{K} \left\{ \frac{1}{\sigma^{[k']}_{t,\ell}} (\mathbf{q}^{[k']})^{\mathsf{H}} \mathbf{z}^{[k']}_{t,\ell} \frac{\partial \phi^{[k]}}{\partial s_{k'}} \left(\left\{ \frac{s^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right\}_k \right) \right. \\ \left. - \frac{1}{\sigma^{[k']}_{t,\ell}} (\mathbf{z}^{[k']}_{t,\ell})^{\mathsf{H}} \mathbf{q}^{[k']} \frac{\partial \phi^{[k]}}{\partial s^*_{k'}} \left(\left\{ \frac{s^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right\}_k \right) \right\} + o_p(N_s^{-1/2}). \tag{60}$$

By the uncorrelatedness of background across mixtures, and by assuming the circularity of $\mathbf{z}^{[k]}_{t,\ell}$, i.e $\mathrm{E}\left[\mathbf{z}^{[k]}_{t,\ell}(\mathbf{z}^{[k]}_{t,\ell})^{\mathsf{T}}\right] = \mathbf{0}$, we can plug (60) into (54) and write

$$\widehat{\boldsymbol{\mu}}^{[k]}_{t,\ell} = \widehat{\mathrm{E}}\left[\phi^{[k]} \left(\left\{ \frac{s^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right\}_k \right) \frac{\mathbf{z}^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right] - \\ - \frac{1}{(\sigma^{[k]}_{t,\ell})^2} \widehat{\mathrm{E}}\left[\frac{\partial \phi^{[k]}}{\partial s_k} \left(\left\{ \frac{s^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right\}_k \right) \mathbf{z}^{[k]}_{t,\ell} (\mathbf{z}^{[k]}_{t,\ell})^{\mathsf{T}} (\mathbf{q}^{[k]})^* \right] - \\ - \frac{1}{(\sigma^{[k]}_{t,\ell})^2} \widehat{\mathrm{E}}\left[\frac{\partial \phi^{[k]}}{\partial s^*_k} \left(\left\{ \frac{s^{[k]}_{t,\ell}}{\sigma^{[k]}_{t,\ell}} \right\}_k \right) \mathbf{z}^{[k]}_{t,\ell} (\mathbf{z}^{[k]}_{t,\ell})^{\mathsf{H}} \mathbf{q}^{[k]} \right] \\ + o_p(N_s^{-1/2}) = \widehat{\boldsymbol{\zeta}}^{[k]}_{t,\ell} - \frac{\widehat{\rho}^{[k]}_{t,\ell} \mathbf{C}^{[k]}_{\mathbf{z},t,\ell} \mathbf{q}^{[k]}}{(\sigma^{[k]}_{t,\ell})^2} + o_p(N_s^{-1/2}). \tag{61}$$

The above results enable us to expand the second term in (49). Now, we turn the expansion of $\widehat{\mathbf{a}}^{[k]}_t$ as a function of $\widehat{\mathbf{w}}^{[k]}$ according to (22). Using (55), we have

$$\widehat{\mathbf{C}}^{[k]}_t \widehat{\mathbf{w}}^{[k]} = \left\langle \begin{bmatrix} (\sigma^{[k]}_{t,\ell})^2 + c^{[k]}_{t,\ell} \\ -\widehat{\boldsymbol{\chi}}^{[k]}_{t,\ell} + \mathbf{C}^{[k]}_{\mathbf{z},t,\ell}\mathbf{q}^{[k]} \end{bmatrix} \right\rangle_\ell + o_p(N_b^{-1/2}), \tag{62}$$

$$(\widehat{\mathbf{w}}^{[k]})^{\mathsf{H}} \widehat{\mathbf{C}}^{[k]}_t \widehat{\mathbf{w}}^{[k]} = \left\langle (\sigma^{[k]}_{t,\ell})^2 + c^{[k]}_{t,\ell} \right\rangle_\ell + o_p(N_b^{-1/2}), \tag{63}$$

and hence

$$\widehat{\mathbf{a}}^{[k]}_t = \begin{bmatrix} 1 \\ \frac{\langle -\widehat{\boldsymbol{\chi}}^{[k]}_{t,\ell} + \mathbf{C}^{[k]}_{\mathbf{z},t,\ell}\mathbf{q}^{[k]} \rangle_\ell}{\langle (\sigma^{[k]}_{t,\ell})^2 \rangle_\ell} \end{bmatrix} + o_p(N_b^{-1/2}). \tag{64}$$

By putting (61) and (64) into (49), we obtain

$$\nabla^{[k]} = \left\langle \widehat{\mathbf{a}}_t^{[k]} - \left[\left\langle - \frac{\widehat{\boldsymbol{\mu}}_{t,\ell}^{[k]}}{\widehat{\nu}_{t,\ell}^{[k]}} \right\rangle_\ell \right] \right\rangle_t$$
$$= \left\langle \left[\begin{array}{c} 0 \\ \frac{\langle -\widehat{\boldsymbol{\chi}}_{t,\ell}^{[k]} + \mathbf{C}_{\mathbf{z},t,\ell}^{[k]} \mathbf{q}^{[k]} \rangle_\ell}{\langle (\sigma_{t,\ell}^{[k]})^2 \rangle_\ell} + \left\langle \frac{\widehat{\boldsymbol{\zeta}}_{t,\ell}^{[k]}}{\widehat{\nu}_{t,\ell}^{[k]}} - \frac{\widehat{\rho}_{t,\ell}^{[k]} \mathbf{C}_{\mathbf{z},t,\ell}^{[k]} \mathbf{q}^{[k]}}{\widehat{\nu}_{t,\ell}^{[k]} (\sigma_{t,\ell}^{[k]})^2} \right\rangle_\ell \end{array} \right] \right\rangle_t +$$
$$+ o_p(N^{-1/2}). \quad (65)$$

The asymptotic solution of $\nabla^{[k]} = \mathbf{0}$ gives us

$$\mathbf{q}^{[k]} = \mathbf{R}_t^{[k]} \left\langle \frac{\langle \widehat{\boldsymbol{\chi}}_{t,\ell}^{[k]} \rangle_\ell}{\langle (\sigma_{t,\ell}^{[k]})^2 \rangle_\ell} - \left\langle \frac{\widehat{\boldsymbol{\zeta}}_{t,\ell}^{[k]}}{\widehat{\nu}_{t,\ell}^{[k]}} \right\rangle_\ell \right\rangle_t + o_p(N^{-1/2}), \quad (66)$$

where

$$\mathbf{R}_t^{[k]} = \left\langle \frac{\langle \mathbf{C}_{\mathbf{z},t,\ell}^{[k]} \rangle_\ell}{\langle (\sigma_{t,\ell}^{[k]})^2 \rangle_\ell} - \left\langle \mathbf{C}_{\mathbf{z},t,\ell}^{[k]} \frac{\rho_{t,\ell}^{[k]}}{(\sigma_{t,\ell}^{[k]})^2 \nu_{t,\ell}^{[k]}} \right\rangle_\ell \right\rangle_t^{-1}. \quad (67)$$

The asymptotic covariance of $\mathbf{q}^{[k]}$ now remains to be derived. For further use, it can be easily shown that

$$\mathrm{E}\left[\langle \widehat{\boldsymbol{\chi}}_{t,\ell}^{[k]} \rangle_\ell \langle \widehat{\boldsymbol{\chi}}_{t,\ell}^{[k]} \rangle_\ell^{\mathsf{H}} \right] = \frac{1}{N_b} \left\langle (\sigma_{t,\ell}^{[k]})^2 \mathbf{C}_{\mathbf{z},t,\ell}^{[k]} \right\rangle_\ell \quad (68a)$$

$$\mathrm{E}\left[\langle \widehat{\boldsymbol{\zeta}}_{t,\ell}^{[k]} \rangle_\ell \langle \widehat{\boldsymbol{\zeta}}_{t,\ell}^{[k]} \rangle_\ell^{\mathsf{H}} \right] = \frac{1}{N_b} \left\langle \frac{\varphi_{t,\ell}^{[k]}}{(\sigma_{t,\ell}^{[k]})^2} \mathbf{C}_{\mathbf{z},t,\ell}^{[k]} \right\rangle_\ell \quad (68b)$$

$$\mathrm{E}\left[\langle \widehat{\boldsymbol{\chi}}_{t,\ell}^{[k]} \rangle_\ell \langle \widehat{\boldsymbol{\zeta}}_{t,\ell}^{[k]} \rangle_\ell^{\mathsf{H}} \right] = \frac{1}{N_b} \left\langle (\nu_{t,\ell}^{[k]})^* \mathbf{C}_{\mathbf{z},t,\ell}^{[k]} \right\rangle_\ell, \quad (68c)$$

where we introduce one more statistic related to the SOI

$$\varphi_{t,\ell}^{[k]} = \mathrm{E}\left[\left| \phi^{[k]} \left(\left\{ \frac{s_{t,\ell}^{[k]}}{\sigma_{t,\ell}^{[k]}} \right\}_k \right) \right|^2 \right]. \quad (69)$$

Using the latter expressions and (66), we arrive at (27), which concludes the proof. ∎

REFERENCES

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

[2] Z. Koldovský, V. Kautský, and P. Tichavský, "Double nonstationarity: Blind extraction of independent nonstationary vector/component from nonstationary mixtures—Algorithms," *IEEE Transactions on Signal Processing*, vol. 70, pp. 5102–5116, 2022.

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

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

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

[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] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, vol. 45, no. 1, pp. 59 – 83, 1995.

[8] 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.

[9] Z. Koldovský, V. Kautský, P. Tichavský, J. Čmejla, and J. Málek, "Dynamic independent component/vector analysis: Time-variant linear mixtures separable by time-invariant beamformers," *IEEE Transactions on Signal Processing*, vol. 69, pp. 2158–2173, 2021.

[10] J. Janský, Z. Koldovský, J. Málek, T. Kounovský, and J. Čmejla, "Auxiliary function-based algorithm for blind extraction of a moving speaker," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2022, no. 1, p. 1, Jan 2022.

[11] V. Kautský, Z. Koldovský, P. Tichavský, and V. Zarzoso, "Cramér–Rao bounds for complex-valued independent component extraction: Determined and piecewise determined mixing models," *IEEE Transactions on Signal Processing*, vol. 68, pp. 5230–5243, 2020.

[12] 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.

[13] M. Anderson, T. Adalı, 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.

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

[15] 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.

[16] D. D. Lee and H. S. Seung, "Learning the parts of objects with nonnegative matrix factorization," *Nature*, vol. 401, pp. 788–791, 1999.

[17] J. Čmejla, Z. Koldovský, V. Kautský, and T. Adalı, "Dynamic independent component extraction with blending mixing vector: Lower bound on mean interference-to-signal ratio," 2022. [Online]. Available: https://arxiv.org/abs/2212.01178

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

[19] 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.

[20] 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.

[21] J.-F. Cardoso, *Likelihood in Handbook of Blind Source Separation*. Academic Press (P. Comon and C. Jutten editors), 2010, ch. 4, pp. 107–154.

[22] G.-S. Fu, R. Phlypo, M. Anderson, and T. Adalı, "Complex independent component analysis using three types of diversity: Non-Gaussianity, nonwhiteness, and noncircularity," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 794–805, 2015.

[23] 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.

[24] 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.

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

[26] V. Kautský, P. Tichavský, Z. Koldovský, and T. Adalı, "Performance bounds for complex-valued independent vector analysis," *IEEE Transactions on Signal Processing*, vol. 68, pp. 4258–4267, 2020.

[27] V. Kautský, Z. Koldovský, and P. Tichavský, "Performance bound for blind extraction of non-Gaussian complex-valued vector component from Gaussian background," in *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, vol. 5287–5291, May 2019.

[28] S. Fortunati, F. Gini, M. S. Greco, and C. D. Richmond, "Performance bounds for parameter estimation under misspecified models: Fundamental findings and applications," *IEEE Signal Processing Magazine*, vol. 34, no. 6, pp. 142–157, 2017.

[29] 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.

[30] 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.
[31] 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.
[32] 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.
[33] 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.
[34] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2011, pp. 189–192.
[35] P. Tichavský and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 878–891, Mar. 2009.
[36] 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.
[37] A. Weiss, S. A. Cheema, M. Haardt, and A. Yeredor, "Performance analysis of the Gaussian quasi-maximum likelihood approach for independent vector analysis," *IEEE Transactions on Signal Processing*, vol. 66, no. 19, pp. 5000–5013, 2018.
[38] M. Anderson, X.-L. Li, and T. Adalı, "Joint blind source separation with multivariate Gaussian model: Algorithms and performance analysis," *IEEE Trans. Signal Processing*, vol. 60, no. 4, pp. 2049–2055, April 2012.
[39] D. Lahat, J.-F. Cardoso, and H. Messer, "Second-order multidimensional ICA: Performance analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4598–4610, 2012.
[40] 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, Sep. 2016.
[41] 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.
[42] G.-S. Fu, R. Phlypo, M. Anderson, and T. Adalı, "Blind source separation by entropy rate minimization," *IEEE Trans. Signal Processing*, vol. 62, no. 16, pp. 4245—4255, Aug. 2014.
[43] 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.
[44] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proceedings of International Conference on Independent Component Analysis and Signal Separation*, Dec. 2001, pp. 722–727.
[45] Z. Koldovský and F. Nesta, "Performance analysis of source image estimators in blind source separation," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4166–4176, Aug. 2017.
[46] Z. Koldovský, J. Málek, and J. Janský, "Extraction of independent vector component from underdetermined mixtures through block-wise determined modeling," in *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, vol. 7903–7907, May 2019.
[47] F. Nesta and Z. Koldovský, "Supervised independent vector analysis through pilot dependent components," in *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, March 2017, pp. 536–540.
[48] A. Brendel, T. Haubner, and W. Kellermann, "A unified probabilistic view on spatially informed source separation and extraction based on independent vector analysis," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3545–3558, 2020.
[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] 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.
[51] P. Tichavsky, Z. Koldovsky, A. Yeredor, G. Gomez-Herrero, and E. Doron, "A hybrid technique for blind separation of non-gaussian and time-correlated sources using a multicomponent approach," *IEEE Transactions on Neural Networks*, vol. 19, no. 3, pp. 421–430, 2008.
[52] H. B. Mann and A. Wald, "On stochastic limit and order relationships," *The Annals of Mathematical Statistics*, vol. 14, no. 3, pp. 217–226, 1943.