

Double Nonstationarity: Blind Extraction of Independent Nonstationary Vector/Component From Nonstationary Mixtures—Algorithms

Zbyněk Koldovský¹, Senior Member, IEEE, Václav Kautský², and Petr Tichavský³, Senior Member, IEEE

Abstract—In this article, nonstationary mixing and source models are combined for developing new fast and accurate algorithms for Independent Component or Vector Extraction (ICE/IVE), one of which stands for a new extension of the well-known FastICA. This model allows for a moving source-of-interest (SOI) whose distribution on short intervals can be (non-)circular (non-)Gaussian. A particular Gaussian source model assuming tridiagonal covariance matrix structures is proposed. It is shown to be beneficial in the frequency-domain speaker extraction problem. The algorithms are verified in simulations. In comparison to the state-of-the-art algorithms, they show superior performance in terms of convergence speed and extraction accuracy.

Index Terms—Blind source extraction, blind source separation, dynamic models, independent component analysis, independent vector analysis, moving sources.

I. INTRODUCTION

A. Topic

Blind Source Separation (BSS) aims at recovering unobserved signals, called sources, from their mixture without additional knowledge [1]. This area has been vital in the signal processing and machine learning communities over the last three decades. It is pertinent to situations where particular information about the sources is missing and only general assumptions can be stated. When the sources are statistically independent, BSS can be solved through Independent Component Analysis (ICA) [2]. Blind Source Extraction (BSE) is a related problem in which the goal is to extract a particular source of interest (SOI). BSE is motivated by the fact that targeting the SOI may often be considerably more cost-effective than separating all of the

sources. A BSE counterpart to ICA is Independent Component Extraction (ICE).

It is also possible to consider multiple mixtures (data-sets) and separate them jointly. Joint BSS (jBSS) is advantageous over BSS in situations when relations/dependencies exist among the sources from different data-sets. Processing one SOI extraction from each mixture jointly, we speak about joint BSE (jBSE). The extension of ICA to jBSS is known as Independent Vector Analysis (IVA); the jBSE counterpart of IVA is Independent Vector Extraction (IVE). This article builds on and contributes to ICE and IVE.

B. State-of-The-Art

The existing ICA/IVA/ICE/IVE algorithms can be categorized based on the statistical model of the sources, referred to as source model, which is used for their development. In general, their goal is to capture various signal features as much as possible. However, mathematical tractability and computational costs must also be taken into account.

In this overview, we focus on two major source models because the key idea of this paper builds on their combination; a survey of BSS methods beyond these classes can be found, e.g., in [3]. 1) The non-Gaussian model considers each source as a sequence of independently and identically distributed (i.i.d.) non-Gaussian random variables. 2) The nonstationary model allows for varying variance and, typically, assumes that sources are sequences of independent Gaussian variables whose variances are changing from interval to interval. The combination of these models occurs when non-Gaussianity is taken into account in the nonstationary model, so sources are assumed to be i.i.d. Gaussian or non-Gaussian within the intervals.

Non-Gaussianity-based ICA methods represent algorithms based on mutual information minimization [2], maximum likelihood estimation (MLE) [4], neural network-like approaches [5], [6], etc.; they were shown to be more or less related with MLE [7]. More advanced methods adapt the source model by matching parametric [8] or non-parametric [9] non-Gaussian distributions to the separated sources. The methods also differ in the optimization approach. For example, there are gradient methods [10], auxiliary-function-based methods [11] or fixed-point algorithms [12].

Non-Gaussianity-based BSE can be accomplished by minimizing the output signal entropy [12]. ICE is based on a reduced

Manuscript received 17 March 2022; revised 5 August 2022; accepted 16 October 2022. Date of publication 25 October 2022; date of current version 7 November 2022. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ali Tajer. This work was supported in part by The Czech Science Foundation under Project 20-17720S, and in part by the Department of the Navy, Office of Naval Research Global, under Project N62909-19-1-2105. (Corresponding author: Zbyněk Koldovský.)

Zbyněk Koldovský and Václav Kautský are with the Acoustic Signal Analysis, Processing Group, Faculty of Mechatronics, Informatics, Interdisciplinary Studies, Technical University of Liberec, 461 17 Liberec, Czech Republic (e-mail: zbynek.koldovsky@tul.cz; kautsvac@fjfi.cvut.cz).

Petr Tichavský is with the The Czech Academy of Sciences, Institute of Information Theory and Automation, 182 08 Prague, Czech Republic (e-mail: tichavsk@utia.cas.cz).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TSP.2022.3216106>, provided by the authors.

Digital Object Identifier 10.1109/TSP.2022.3216106

mixing model parameterization, in which one source is treated as the SOI and the others, which are not subject to separation, as background sources [13]. ICE has been shown equivalent with the minimum entropy approach when the background model is multivariate Gaussian [14] and with ICA when it is multivariate non-Gaussian [15], [16].

In contrast to the non-Gaussian model, the nonstationarity-based one can be identified using second-order statistics (SOS) only [17], [18]. Numerous methods are based on the Joint Approximate Diagonalization (JAD) of sample covariance matrices computed on intervals (blocks) of data [18], [19], [20]. For BSE, JAD can be replaced by Joint Block Diagonalization (JBD) where the SOI is represented by one-dimensional subspace that is separated from a hyperplane representing the background [21], [22], [23].

Similarly, the source models have been applied in jBSS and jBSE to model vector sources, where a vector source consists of corresponding scalar sources, one source per mixture (dataset). In IVA and IVE, this approach entails using multivariate non-Gaussian model distributions that capture internal dependencies among the scalar sources [13], [24], [25], [26], [27]. The nonstationary model can be effectively used for jBSS when the elements of vector sources are correlated [28], [29]; see also [23].

The non-Gaussian and nonstationary source models have been successfully combined in ICA [30], [31], [32] as well as in IVA [33]. The recent extensions of IVA known under the umbrella of Independent Low Rank Matrix Analysis (ILRMA) can also be considered as extensions of this kind [34], [35], [36].

Another classification of BSS methods is based on the assumed model of source mixing, that is, the mixing model. In the vast majority of BSS literature, the instantaneous linear mixing model is assumed [1], [2], [37], [38]. The other intensively studied convolutive model is also linear. It is often considered in the Fourier transform domain where it is translated to a set of instantaneous mixtures, which can be treated as the jBSS problem [39], [40], [41]. Some specific nonlinear mixing models have been studied, e.g., in [42], [43].

Similarly to source models, nonstationarity can be brought into the mixing models¹. The goal is to capture the time-variant mixing conditions caused, e.g., by source movements or similar changes. Typically, estimation methods for the static linear mixing are turned into adaptive algorithms [44], [45]. The nonstationary mixing process is less frequently described by a more specific parameterization such as that used in [46], [47]. Recently, semi-time-variant models denoted as CMV and CSV (Constant Mixing/Separating Vector) have been considered in [14], [16], [48]. CMV and CSV are designed for BSE/jBSE in which the SOI is static or moving, respectively, on a dynamic background. The nonstationarity is arranged through allowing specific parameters to be changing from interval to interval. In [14], the well-known FastICA algorithm [12] has been extended for CSV and named as FastDIVA (Fast Dynamic IVA).

C. Contribution

Although FastDIVA builds on the non-Gaussian model, it partly allows for source nonstationarity. This is because the

¹Nonstationary mixing models are sometimes termed as “dynamic models” [14].

variance of the SOI is allowed to change over the intervals of the CSV mixing model. However, the performance analysis of FastDIVA in [14] as well as the Cramér-Rao bound in [16] have shown that the SOI is not identifiable when its distribution is Gaussian. Dividing data into more (shorter) intervals does not seem very effective since the number of mixing parameters proportionally grows, and, moreover, the Gaussian SOI remains unidentifiable.

This is the main motivation behind the novel extension provided in this article: We propose to combine the nonstationary CSV mixing model with the nonstationary source model, enabling Gaussian and/or non-Gaussian moving SOI. Two second-order algorithms are derived, one of which stands for a new extension of FastDIVA (resp. FastICA and FastIVA). Special attention is given to the Gaussian source model, for which distinguished SOS-based variants of the algorithms are derived. The latter can also efficiently benefit from the SOI non-circularity. Moreover, a particular Gaussian source model for the SOI, which assumes tridiagonal covariance matrix structures, is proposed. This model is efficiently implemented within the proposed algorithms, and it shows promising results in the frequency-domain speaker extraction problem where the K -dimension (the number of frequencies/mixtures/data-sets) can take value in the order of hundreds. These methods are verified by extensive numerical studies. In comparison to state-of-the-art algorithms, they show superior performance in terms of convergence speed and extraction accuracy.

The paper is organized as follows. Technical description of the problem is formulated in Section II. In Section III, the second-order algorithms are derived. Section IV is devoted to special variants of the algorithms based on the Gaussian source model. Experimental validation is provided in Section V; and Section VI concludes the paper.

D. Nomenclature and conventions

Plain, bold, and bold capital letters denote scalars, vectors, and matrices, respectively. Upper index \cdot^T , \cdot^H , or \cdot^* denotes, respectively, transposition, conjugate transpose, or complex conjugate. The Matlab convention for matrix/vector concatenation will be used, e.g., $[1; \mathbf{g}] = [1, \mathbf{g}^T]^T$. $\mathbb{E}[\cdot]$ stands for the expectation value of the argument, and $\hat{\mathbb{E}}[\cdot]$ is the average value of the argument taken over all of its available samples. Throughout this article, the index of a data-set, block, and sub-block, will always be denoted, respectively, by $k = 1, \dots, K$, $t = 1, \dots, T$, and $\ell = 1, \dots, L$. $\{\cdot\}_k$ is a short notation of the argument with all values of index k , e.g., $\{\mathbf{w}_k\}_k$ means $\mathbf{w}_1, \dots, \mathbf{w}_K$. The average value of a_t taken over all available blocks, i.e., $\frac{1}{T} \sum_{t=1}^T a_t$, is denoted by $\langle a_t \rangle_t$; similarly, $\langle a_\ell \rangle_\ell$ denotes the average taken over all sub-blocks, i.e., $\frac{1}{L} \sum_{\ell=1}^L a_\ell$.

II. PROBLEM FORMULATION

We consider measurements of length N in K data-sets, each one obtained by d sensors, i.e., of dimension d . Each measurement is divided into T non-overlapping intervals of length N_b , hereafter called blocks, and each block is divided into L sub-blocks of length N_s . For simplicity, the blocks (and sub-blocks) have the same length, although different lengths

could be considered as well. Hence, $N = T \cdot N_b$, $N_b = L \cdot N_s$, and $N = T \cdot L \cdot N_s$. All signals and parameters are, in general, complex-valued.

A. Semi-Time-Variant Mixing Model

The n th sample of the measured data, $n = 1, \dots, N_s$, within the ℓ th sub-block of the t th block and in the k th data-set is modeled as linear instantaneous mixture

$$\mathbf{x}_{k,t,\ell}(n) = \mathbf{A}_{k,t} \mathbf{u}_{k,t,\ell}(n), \quad (1)$$

where $\mathbf{A}_{k,t}$ is the $d \times d$ mixing matrix. The signal samples will be assumed i.i.d. within the sub-blocks; therefore, the argument n can be omitted; $\mathbf{u}_{k,t,\ell} = [s_{k,t,\ell}; \mathbf{z}_{k,t,\ell}]$ is the vector of the source components $s_{k,t,\ell}$ and $\mathbf{z}_{k,t,\ell}$ representing, respectively, the SOI and background. The fundamental assumption of ICE/IVE is that $s_{k,t,\ell}$ and $\mathbf{z}_{k,t,\ell}$ are independent random variables; let their mean value be zero.

Since the last $d - 1$ columns of the mixing matrix are irrelevant to the SOI extraction problem (only their implied subspace is relevant), we can consider a parameterization of $\mathbf{A}_{k,t}$ in which its last $d - 1$ columns span the same subspace. It was shown in [13] that $\mathbf{A}_{k,t}$ can be thus parameterized as²

$$\mathbf{A}_{k,t} = \begin{pmatrix} \mathbf{a}_{k,t} & \mathbf{Q}_{k,t} \end{pmatrix} = \begin{pmatrix} \gamma_{k,t} & \mathbf{h}_k^H \\ \mathbf{g}_{k,t} & \frac{1}{\gamma_{k,t}} (\mathbf{g}_{k,t} \mathbf{h}_k^H - \mathbf{I}_{d-1}) \end{pmatrix}. \quad (2)$$

$\mathbf{a}_{k,t}$ is called the mixing vector (the first column of $\mathbf{A}_{k,t}$) corresponding to the SOI; \mathbf{I}_d denotes the $d \times d$ identity matrix. Note that (1) can also be written in the form

$$\mathbf{x}_{k,t,\ell} = \mathbf{a}_{k,t} s_{k,t,\ell} + \mathbf{y}_{k,t,\ell}, \quad (3)$$

where $\mathbf{y}_{k,t,\ell} = \mathbf{Q}_{k,t} \mathbf{z}_{k,t,\ell}$ play the role of the background signals as they are observed by the sensors.

Equivalently to (1), the de-mixing model reads

$$\mathbf{u}_{k,t,\ell} = \mathbf{W}_{k,t} \mathbf{x}_{k,t,\ell}, \quad (4)$$

where

$$\mathbf{W}_{k,t} = \begin{pmatrix} \mathbf{w}_k^H \\ \mathbf{B}_{k,t} \end{pmatrix} = \begin{pmatrix} \beta_k^* & \mathbf{h}_k^H \\ \mathbf{g}_{k,t} & -\gamma_{k,t} \mathbf{I}_{d-1} \end{pmatrix}, \quad (5)$$

where $\mathbf{w}_k = [\beta_k; \mathbf{h}_k]$ is the separating vector that satisfies the distortionless constraint $\mathbf{w}_k^H \mathbf{a}_{k,t} = 1$, $t = 1, \dots, T$. Under this constraint, the reader can easily verify that $\mathbf{W}_{k,t}$ in (5) is the inverse matrix of $\mathbf{A}_{k,t}$ in (2). It holds that $\det \mathbf{W}_{k,t} = (-1)^{d-1} \gamma_{k,t}^{d-2}$; see Eq. (15) in [13]. $\mathbf{B}_{k,t} = [\mathbf{g}_{k,t} \quad -\gamma_{k,t} \mathbf{I}_{d-1}]$, which depends purely on the mixing vector $\mathbf{a}_{k,t}$, is called the blocking matrix.

Note that the separating vectors \mathbf{w}_k are chosen as independent of t while the mixing vectors $\mathbf{a}_{k,t}$ depend on it. This parameterization corresponds to the semi-time-variant CSV mixing model advocated in [14], [16]. The varying $\mathbf{a}_{k,t}$ entail a changing

²In (2), $\gamma_{k,t}$ appears in a denominator, which calls for rows permutation in case that $\gamma_{k,t} = 0$. We therefore assume that $\gamma_{k,t} \neq 0$, nevertheless, as will be seen later in the article, $\gamma_{k,t}$ does not appear in a denominator within the update rules of the derived algorithms. So this assumption does not bring any practical restriction.

location of the SOI (i.e., movements), while the constant \mathbf{w}_k ensures the existence of a beamformer that extracts the SOI from all the locations; see [48] for the application of this model in the moving speaker extraction problem.

In the sequel, $\mathbf{s}_{t,\ell} = (s_{1,t,\ell}, \dots, s_{K,t,\ell})^T$ will refer to the vector component of the SOI; $s_{k,t,\ell}$ and $\mathbf{z}_{k,t,\ell}$ will be, respectively, called the k th component of the SOI and of the background.

B. Source Model

It is worth pointing out that the mixing parameters remain constant within the blocks while the source model is i.i.d. (stationary) only within the sub-blocks. This means that signals are allowed to be more dynamic than the changes in the mixing process; this approach finds application in many real situations.

As for the SOI, $s_{1,t,\ell}, \dots, s_{K,t,\ell}$ are modeled jointly, and their the joint probability density function (pdf) is denoted by $p_{t,\ell}(\mathbf{s}_{t,\ell})$. The idea of joint statistical modeling is adopted from IVA. It allows for mutual dependencies among the components of the SOI, which helps in solving the permutation problem [49]. Our extension here is that the pdf is allowed to vary across blocks and sub-blocks (dependent on t and ℓ). Since $p_{t,\ell}(\mathbf{s}_{t,\ell})$ is not known, it was proposed in [14] that an appropriate surrogate is

$$p_{t,\ell}(\mathbf{s}_{t,\ell}) \approx f \left(\left\{ \frac{s_{k,t,\ell}}{\hat{\sigma}_{k,t,\ell}} \right\}_k \right) \left(\prod_{k=1}^K \hat{\sigma}_{k,t,\ell} \right)^{-2}, \quad (6)$$

where $f(\cdot)$ is a suitable normalized multivariate pdf³, and $\hat{\sigma}_{k,t,\ell}^2$ is the sample-based variance of the estimate of $s_{k,t,\ell}$.

For the background probabilistic model we assume that all background signals $\mathbf{z}_{k,t,\ell}$ are jointly circular Gaussian, with $\mathbf{z}_{k,t,\ell} \sim \mathcal{CN}(\mathbf{0}, \mathbf{C}_z^{k,t,\ell})$ where $\mathbf{C}_z^{k,t,\ell}$ is an unknown covariance matrix. The background signals from different data-sets are assumed uncorrelated (hence, owing to the Gaussianity, also independent).

The fact that the non-Gaussianity, non-circularity, and the dependencies among the components of the background signals are not assumed, brings about important simplifications into algorithms and bounds derivations. As it has been observed with similar problems, the probability model mismatch does not usually cause algorithm malfunction. Typically, the price for the simplification is a suboptimality in terms of the theoretical achievable extraction accuracy [13], [15].

C. Contrast Function

The contrast function is a function of the mixing parameters whose optimum points provide their consistent estimates. The function is derived from the likelihood function by replacing unknown pdfs and nuisance parameters; it is sometimes referred to as the quasi-likelihood function [4], [29].

By comparing the mixing and source models with the one in [14], the model presented here differs only in that $\sigma_{k,t,\ell}^2$ and

³The model density $f(\cdot)$ could have been considered as dependent on t and ℓ . However, since there is typically lack of information about the pdf of the SOI, we find it more practical when $f(\cdot)$ is constant and the variability of the pdf is captured only by the time-varying variance $\hat{\sigma}_{k,t,\ell}^2$.

$\mathbf{C}_z^{k,t,\ell}$ are allowed to be changing over the sub-blocks. This allows us to obtain the contrast function by straightforward modifications of Eq. (12) in [14].

Namely, the signals and their parameters become dependent on the sub-block index ℓ . Therefore, the sample-average operator denoted by $\hat{\mathbb{E}}[\cdot]$ computes the average only over the samples in the sub-block (not over the block as in [14]). The mixing parameters remain the same, i.e., independent of ℓ . Finally, the whole formula must be averaged over the sub-blocks. Therefore, the contrast function takes on the form

$$\begin{aligned} \mathcal{C}(\{\mathbf{w}_k, \mathbf{a}_{k,t}\}_{k,t}) = & \left\langle \left\langle \hat{\mathbb{E}} \left[\log f \left(\left\{ \frac{\hat{s}_{k,t,\ell}}{\hat{\sigma}_{k,t,\ell}} \right\}_k \right) \right] \right. \right. \\ & \left. \left. - \sum_{k=1}^K \log \hat{\sigma}_{k,t,\ell}^2 - \sum_{k=1}^K \hat{\mathbb{E}} \left[\hat{\mathbf{z}}_{k,t,\ell}^H (\mathbf{C}_z^{k,t,\ell})^{-1} \hat{\mathbf{z}}_{k,t,\ell} \right] \right\rangle \right\rangle_{\ell} \\ & + (d-2) \sum_{k=1}^K \left\langle \log |\gamma_{k,t}|^2 \right\rangle_t, \end{aligned} \quad (7)$$

where $\hat{s}_{k,t,\ell} = \mathbf{w}_k^H \mathbf{x}_{k,t,\ell}$ is the estimate of the SOI, $\hat{\mathbf{z}}_{k,t,\ell} = \mathbf{B}_{k,t} \mathbf{x}_{k,t,\ell}$ is the estimated background, and $\hat{\sigma}_{k,t,\ell}^2$ denotes the sample-based variance of $\hat{s}_{k,t,\ell}$. The operators $\langle \cdot \rangle_t$ and $\langle \cdot \rangle_{\ell}$ denote averaging over the index t and ℓ , respectively.

D. Indeterminacies

In the (j)BSE, there are two major inherent indeterminacies of the SOI. First, the scales of $s_{k,t,\ell}$ and of $\mathbf{a}_{k,t}$ are ambiguous in the sense that they can be replaced, respectively, by $\alpha_{k,t} s_{k,t,\ell}$ and $\alpha_{k,t}^{-1} \mathbf{a}_{k,t}$ where $\alpha_{k,t} \neq 0$. The scaling ambiguity is, in fact, taken into account in (6) through the consideration of the normalized pdf. Practical issues due to the scaling ambiguity can be solved through the reconstruction of the SOI's contributions on sensors (so-called images); see [50], [51].

Second, more importantly, there is the fact that the role of the SOI is interchangeable with any independent source in the mixture. Therefore, the contrast function (7) is, in general, not convex/concave and can have several local extremes, one of which corresponds to the SOI. There is no way to distinguish the SOI from the other sources without additional information, which is unavailable in the blind setting considered in this work. Therefore, throughout the article, we assume (in theoretical developments and in simulations) that a given algorithm is initialized in the domain of attraction of the SOI. The problem of controlling BSE as well as other source extraction algorithms to converge to the desired source(s) has been an important topic in various applications; see e.g. [52], [53], [54].

III. PROPOSED ALGORITHMS

In this section, we derive second-order derivative-based algorithms seeking for the desired optimum point of the contrast function (7). For their development, we make use of the fact that the terms in (7) are mostly separated. Therefore, we simplify the exposition as if $K = 1$ and $T = 1$; so the indices k and t can be dropped. The extension to $K > 1$ and $T > 1$ will be discussed later.

For $K = 1$ and $T = 1$, we have the time-invariant instantaneous one-mixture problem studied under the umbrella of ICE [13]. The contrast function is simplified to

$$\begin{aligned} \mathcal{C}_1(\mathbf{w}, \mathbf{a}) = & \left\langle \hat{\mathbb{E}} \left[\log f \left(\frac{\hat{s}_{\ell}}{\hat{\sigma}_{\ell}} \right) \right] - \log \hat{\sigma}_{\ell}^2 - \hat{\mathbb{E}} \left[\hat{\mathbf{z}}_{\ell}^H \mathbf{R}_{\ell} \hat{\mathbf{z}}_{\ell} \right] \right\rangle_{\ell} \\ & + (d-2) \log |\gamma|^2, \end{aligned} \quad (8)$$

where we have introduced auxiliary matrices \mathbf{R}_{ℓ} , whose ideal value is $\mathbf{R}_{\ell} = (\mathbf{C}_z^{\ell})^{-1}$. Since \mathbf{C}_z^{ℓ} are not known, we select the value of \mathbf{R}_{ℓ} later in Lemma 1.

A. Orthogonal Constraint

The parameter vectors \mathbf{w} and \mathbf{a} are almost free, linked only through the distortionless constraint $\mathbf{w}^H \mathbf{a} = 1$. Since the contrast function has many spurious extremes where \mathbf{w} and \mathbf{a} do not correspond to the same source, it is helpful to link \mathbf{w} and \mathbf{a} more tightly using the orthogonal constraint (OGC).

By definition, the OGC requires that the sample correlations of the estimated SOI and background be zero. When imposing the OGC for each sub-block, that is,

$$\hat{\mathbb{E}}[\hat{s}_{\ell}^* \hat{\mathbf{z}}_{\ell}] = 0, \quad \ell = 1, \dots, L, \quad (9)$$

we have $L(d-1)$ conditions, which, together with the distortionless constraint $\mathbf{w}^H \mathbf{a} = 1$, provides $L(d-1) + 1$ linear conditions on \mathbf{a} (or \mathbf{w}) in total. They cannot in general be satisfied all simultaneously unless $L = 1$.

Since the case $L > 1$ is of particular interest in this work, we propose to replace (9) by a weaker condition

$$\left\langle \hat{\mathbb{E}}[\hat{s}_{\ell}^* \hat{\mathbf{z}}_{\ell}] \right\rangle_{\ell} = 0, \quad (10)$$

which imposes orthogonality over the whole block of signals (ignoring sub-blocks).⁴ When \mathbf{a} is treated as the dependent variable, we can apply the formula derived in Appendix A in [13], and the solution of (10) satisfying $\mathbf{w}^H \mathbf{a} = 1$ is

$$\mathbf{a} = \frac{\hat{\mathbf{C}} \mathbf{w}}{\mathbf{w}^H \hat{\mathbf{C}} \mathbf{w}}, \quad (11)$$

where $\hat{\mathbf{C}} = \left\langle \hat{\mathbf{C}}_{\ell} \right\rangle_{\ell}$ and $\hat{\mathbf{C}}_{\ell} = \hat{\mathbb{E}}[\mathbf{x}_{\ell} \mathbf{x}_{\ell}^H]$ is the sample covariance matrix of \mathbf{x}_{ℓ} .

B. Gradient

The first step for deriving the algorithms is to compute the gradient of (8) with respect to \mathbf{w}^H when \mathbf{a} is dependent through (11). We summarize the result in the following Lemma.

Lemma 1: Let, after computing the derivatives, the matrices \mathbf{R}_{ℓ} be put equal to $\mathbf{R}_{\ell} = \left\langle \hat{\mathbf{C}}_{\ell}^{-1} \right\rangle_{\ell}$ where $\hat{\mathbf{C}}_{\ell} = \hat{\mathbb{E}}[\hat{\mathbf{z}}_{\ell} \hat{\mathbf{z}}_{\ell}^H]$ is the sample-based covariance matrix of \mathbf{z}_{ℓ} . It then holds that

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}^H} \mathcal{C}_1 \left(\mathbf{w}, \frac{\hat{\mathbf{C}} \mathbf{w}}{\mathbf{w}^H \hat{\mathbf{C}} \mathbf{w}} \right) = \\ \mathbf{a} - \left\langle \hat{\mathbb{E}} \left[\phi \left(\frac{\hat{s}_{\ell}}{\hat{\sigma}_{\ell}} \right) \frac{\mathbf{x}_{\ell}}{\hat{\sigma}_{\ell}} \right] + \Re(\hat{\nu}_{\ell}) \mathbf{a}_{\ell} - \mathbf{a}_{\ell} \right\rangle_{\ell}, \end{aligned} \quad (12)$$

⁴A reviewer of this article rightly noted that an appropriately weighted alternative to the mean could be considered in the future.

where $\Re(\cdot)$ denotes the real part of the argument,

$$\phi(s) = -\frac{\partial}{\partial s} \log f(s) \quad (13)$$

is the score function of the model density $f(s)$, $\hat{\nu}_\ell$ is the sample-based estimate of

$$\nu_\ell = \mathbb{E} \left[\phi \left(\frac{s_\ell}{\sigma_\ell} \right) \frac{s_\ell}{\sigma_\ell} \right], \quad (14)$$

and

$$\mathbf{a}_\ell = \frac{\hat{\mathbf{C}}_\ell \mathbf{w}}{\mathbf{w}^H \hat{\mathbf{C}}_\ell \mathbf{w}}. \quad (15)$$

Proof: See Appendix A. \blacksquare

We now need to make an adjustment of the model density $f(\cdot)$ for the sake of consistency. The problem is revealed by the following Lemma. Hereafter, \mathbf{w}^* will denote the true separating vector such that $(\mathbf{w}^*)^H \mathbf{x}_\ell = s_\ell$, $\ell = 1, \dots, L$.

Lemma 2: Let $\mathbf{w} = \mathbf{w}^*$ and $N \rightarrow +\infty$. Then, the right-hand side of (12) converges to

$$(2 - \langle \Re(\nu_\ell) + \nu_\ell \rangle_\ell) \mathbf{a}. \quad (16)$$

Proof: For $N \rightarrow +\infty$, it holds that $\mathbf{a}_\ell \rightarrow \mathbf{a}$, $\hat{\nu}_\ell \rightarrow \nu_\ell$, and by (3) and using the fact that s_ℓ and \mathbf{y}_ℓ are independent,

$$\hat{\mathbb{E}} \left[\phi \left(\frac{s_\ell}{\hat{\sigma}_\ell} \right) \frac{\mathbf{x}_\ell}{\hat{\sigma}_\ell} \right] \xrightarrow{N \rightarrow +\infty} \mathbb{E} \left[\phi \left(\frac{s_\ell}{\sigma_\ell} \right) \frac{\mathbf{a}s_\ell + \mathbf{y}_\ell}{\sigma_\ell} \right] = \nu_\ell \mathbf{a}. \quad (17)$$

The assertion of the lemma follows. \blacksquare

To make the stationary point of the contrast function a consistent estimate of \mathbf{w} , (16) must be equal to zero when $\mathbf{w} = \mathbf{w}^*$ and $N \rightarrow +\infty$. As observed in previous works [13], [14], this problem appears due to the arbitrarily chosen model density $f(\cdot)$. It can easily be solved by considering a suitable sub-block-dependent modification of $f(\cdot)$. This step is performed through the substitution $\phi(\cdot) \rightarrow \hat{\nu}_\ell^{-1} \phi(\cdot)$. Consequently, the gradient (12) turns to

$$\nabla = \mathbf{a} - \left\langle \hat{\nu}_\ell^{-1} \hat{\mathbb{E}} \left[\phi \left(\frac{\hat{s}_\ell}{\hat{\sigma}_\ell} \right) \frac{\mathbf{x}_\ell}{\hat{\sigma}_\ell} \right] \right\rangle_\ell. \quad (18)$$

The reader can verify that, for $\mathbf{w} = \mathbf{w}^*$ and $N \rightarrow +\infty$, $\nabla = \mathbf{0}$ holds, which ensures consistency.

C. Hessian

The Hessian matrices of the real-valued contrast function defined using the Wirtinger calculus are given by [55]

$$\mathbf{H}_1^{\text{expl}} = \frac{\partial^2 \mathcal{C}_1}{\partial \mathbf{w}^T \partial \mathbf{w}} = \frac{\partial \nabla^H}{\partial \mathbf{w}}, \quad (19)$$

$$\mathbf{H}_2^{\text{expl}} = \frac{\partial^2 \mathcal{C}_1}{\partial \mathbf{w}^H \partial \mathbf{w}} = \frac{\partial \nabla^T}{\partial \mathbf{w}}. \quad (20)$$

The superscript expl is used to distinguish the explicit Hessian matrices from their counterparts that are finally used in the algorithms. These are obtained by considering the analytical shapes of $\mathbf{H}_1^{\text{expl}}$ and $\mathbf{H}_2^{\text{expl}}$ when $\mathbf{w} = \mathbf{w}^*$ and $N \rightarrow +\infty$.

We consider two approaches: the derivatives of (18) are considered with and without the imposed OGC on \mathbf{a} , respectively.

In both computations, the $\hat{\nu}_\ell s$ variables are treated as constants. The results are summarized in the following two Lemmas.

Lemma 3: Let $\mathbf{w} = \mathbf{w}^*$, $N \rightarrow +\infty$, $\hat{\nu}_\ell$ be constants, and \mathbf{a} depend on \mathbf{w} through (11). It then holds that $\mathbf{H}_1^{\text{expl}} \rightarrow \mathbf{H}_1$ and $\mathbf{H}_2^{\text{expl}} \rightarrow \mathbf{H}_2$, where

$$\mathbf{H}_1^* = \left\langle \nu_\ell^{-1} \left(\frac{\tau_\ell}{2} \mathbf{a}_\ell - (\nu_\ell + \eta_\ell) \mathbf{a} \right) \right\rangle_\ell \mathbf{a}^T, \quad (21)$$

$$\mathbf{H}_2 = \frac{\langle \mathbf{C}_\ell^* \rangle_\ell}{\langle \sigma_\ell^2 \rangle_\ell} - \left\langle \frac{\rho_\ell \mathbf{C}_\ell^*}{\nu_\ell \sigma_\ell^2} \right\rangle_\ell - \left\langle \nu_\ell^{-1} \left(\omega_\ell \mathbf{a}^* - \frac{\tau_\ell}{2} \mathbf{a}_\ell^* \right) \right\rangle_\ell \mathbf{a}^T, \quad (22)$$

where $\tau_\ell = \eta_\ell + \xi_\ell + \nu_\ell$ and $\omega_\ell = \xi_\ell + \nu_\ell - \rho_\ell$, and

$$\rho_\ell = \mathbb{E} \left[\frac{\partial \phi \left(\frac{s_\ell}{\sigma_\ell} \right)}{\partial s^*} \right], \quad (23)$$

$$\xi_\ell = \mathbb{E} \left[\frac{\partial \phi \left(\frac{s_\ell}{\sigma_\ell} \right) |s_\ell|^2}{\partial s^* \sigma_\ell^2} \right], \quad (24)$$

$$\eta_\ell = \mathbb{E} \left[\frac{\partial \phi \left(\frac{s_\ell}{\sigma_\ell} \right) s_\ell^2}{\partial s \sigma_\ell^2} \right]. \quad (25)$$

Proof: See Appendix B. \blacksquare

Lemma 4: Let $\mathbf{w} = \mathbf{w}^*$, $N \rightarrow +\infty$, and $\hat{\nu}_\ell$ and \mathbf{a} be constants. It then holds that $\mathbf{H}_1^{\text{expl}} \rightarrow \mathbf{H}_1$ and $\mathbf{H}_2^{\text{expl}} \rightarrow \mathbf{H}_2$, where

$$\mathbf{H}_1^* = \left\langle \nu_\ell^{-1} \left(\frac{\tau_\ell}{2} \mathbf{a}_\ell - \eta_\ell \mathbf{a} \right) \right\rangle_\ell \mathbf{a}^T, \quad (26)$$

$$\mathbf{H}_2 = - \left\langle \frac{\rho_\ell \mathbf{C}_\ell^*}{\nu_\ell \sigma_\ell^2} \right\rangle_\ell \nu_\ell^{-1} \left((\xi_\ell - \rho_\ell) \mathbf{a}^* - \frac{\tau_\ell}{2} \mathbf{a}_\ell^* \right) \mathbf{a}^T. \quad (27)$$

Proof: See Appendix B. \blacksquare

D. Learning Rule

The learning rule in the proposed algorithms is inspired by the exact Newton-Raphson (NR) update derived in [55]. An iteration of the exact NR algorithm is given by

$$\mathbf{w}^{\text{new}} = \mathbf{w} - \mathbf{H}_3^{-1} (\nabla - (\mathbf{H}_1^{\text{expl}})^* (\mathbf{H}_2^{\text{expl}})^{-1} \nabla^*), \quad (28)$$

where $\mathbf{H}_3 = (\mathbf{H}_2^{\text{expl}})^* - (\mathbf{H}_1^{\text{expl}})^* (\mathbf{H}_2^{\text{expl}})^{-1} \mathbf{H}_1^{\text{expl}}$. We employ this update with the following two modifications:

- 1) $\mathbf{H}_1^{\text{expl}}$ and $\mathbf{H}_2^{\text{expl}}$ are replaced, respectively, by \mathbf{H}_1 and \mathbf{H}_2 , in which the unknown signal statistics ν_ℓ , ρ_ℓ , ... are replaced by their sample-based estimates using samples of the current estimate of the SOI, and
- 2) the rank-one terms in \mathbf{H}_1 and \mathbf{H}_2 are neglected, hence, the entire \mathbf{H}_1 is put equal to zero.

After these modifications, the update rule (28) is simplified to

$$\mathbf{w}^{\text{new}} = \mathbf{w} - \mathbf{H}^{-1} \nabla, \quad (29)$$

where ∇ is computed the same as in (18) and

$$\mathbf{H} = \frac{\langle \hat{\mathbf{C}}_\ell \rangle_\ell}{\langle \hat{\sigma}_\ell^2 \rangle_\ell} - \left\langle \frac{\hat{\rho}_\ell \hat{\mathbf{C}}_\ell}{\hat{\nu}_\ell^* \hat{\sigma}_\ell^2} \right\rangle_\ell \quad (30)$$

Algorithm 1: General Scheme of the Proposed Algorithms for Blind Source Extraction.

Input: \mathbf{x} , \mathbf{w}_{ini} , tol

Output: \mathbf{a} , \mathbf{w}

```

1  $\mathbf{w} = \mathbf{w}_{\text{ini}}$ 
2 repeat
3    $\mathbf{w}_{\text{old}} = \mathbf{w}$ ;
4   Update  $\mathbf{a}$  according to (11);
5   Update  $\mathbf{w}$  according to (29);
6    $\text{crit} = 1 - \frac{|\mathbf{w}^H \mathbf{w}_{\text{old}}|}{\|\mathbf{w}\| \|\mathbf{w}_{\text{old}}\|}$ ;
7 until  $\text{crit} < \text{tol}$ ;
```

for the first proposed algorithm based on Lemma 3, and

$$\mathbf{H} = - \left\langle \frac{\hat{\rho}_\ell \hat{\mathbf{C}}_\ell}{\hat{\nu}_\ell^* \hat{\sigma}_\ell^2} \right\rangle_\ell \quad (31)$$

for the second proposed algorithm based on Lemma 4. Both algorithms are iterated according to the scheme given by Algorithm III-D until convergence prevails, where the same stopping rule (line 6 in Algorithm III-D) is used as that in [12]. For the sake of consistency with our previous works [14], [56], the algorithms will be referred to as FastDIVA and QuickIVE, respectively.

Our neglecting the rank-1 terms in \mathbf{H}_1 and \mathbf{H}_2 is justified by the fact that, in our experiments, we did not observe any practical improvement when these terms were kept. Similar simplification has been used in [14] when $L = T = 1$, where it is justified by Proposition 2 (of [14]) in a mathematically rigorous way.

E. Extension to $T > 1$ and $K > 1$

We now can get back to the original notation with all three indices k , t , and ℓ (data-set, block, and sub-block) and admit that $T > 1$ and $K > 1$. The extension of the results derived in Sections III-B and III-C to $T > 1$ is straightforward. The gradient and the Hessian matrices are averaged over t due to the averaging operator in (7).

For the extension to $K > 1$, we use the fact that the parameters in (7) with different k are separated (depend only on the parameters with the same k) up to the first term involving $f(\cdot)$, which is a function of the entire vector component of the SOI, hence, of all the separating vectors $\mathbf{w}_1, \dots, \mathbf{w}_K$. Here, we simplify the algorithm development by ignoring mixed second-order derivatives⁵, i.e., the derivatives by \mathbf{w}_{k_1} and \mathbf{w}_{k_2} , $k_1 \neq k_2$. This is equivalent to assuming that the Hessian matrices related to the concatenated parameter vector $\mathbf{w} = [\mathbf{w}_1; \dots; \mathbf{w}_K]$ are block-diagonal with $K \times K$ blocks on the main diagonal where the k th block corresponds to the second order derivatives by \mathbf{w}_k .

Consequently, the algorithm derivations done in the previous sections apply separately to each $k = 1, \dots, K$, and we only need to extend the definitions (14), (23)-(25) as follows. Let the

k th score function related to $f(\cdot)$ be defined as

$$\phi_k(\mathbf{s}_{t,\ell}) = - \frac{\partial}{\partial s_k} \log f(\mathbf{s}_{t,\ell}), \quad (32)$$

where the partial derivative is taken over the k th argument denoted by s_k . We can now continue with the new definitions

$$\nu_{k,t,\ell} = \mathbb{E} \left[\phi_k \left(\left\{ \frac{s_{k,t,\ell}}{\sigma_{k,t,\ell}} \right\}_k \right) \frac{s_{k,t,\ell}}{\sigma_{k,t,\ell}} \right], \quad (33)$$

$$\rho_{k,t,\ell} = \mathbb{E} \left[\frac{\partial \phi_k \left(\left\{ \frac{s_{k,t,\ell}}{\sigma_{k,t,\ell}} \right\}_k \right)}{\partial s_k^*} \right], \quad (34)$$

$$\xi_{k,t,\ell} = \mathbb{E} \left[\frac{\partial \phi_k \left(\left\{ \frac{s_{k,t,\ell}}{\sigma_{k,t,\ell}} \right\}_k \right)}{\partial s_k^*} \frac{|s_{k,t,\ell}|^2}{\sigma_{k,t,\ell}^2} \right], \quad (35)$$

$$\eta_{k,t,\ell} = \mathbb{E} \left[\frac{\partial \phi_k \left(\left\{ \frac{s_{k,t,\ell}}{\sigma_{k,t,\ell}} \right\}_k \right)}{\partial s_k} \frac{s_{k,t,\ell}^2}{\sigma_{k,t,\ell}^2} \right]. \quad (36)$$

The update rules for $K > 1$ and $T > 1$ are given by

$$\mathbf{w}_k^{\text{new}} = \mathbf{w}_k - \mathbf{H}_k^{-1} \nabla_k, \quad k = 1, \dots, K, \quad (37)$$

where

$$\nabla_k = \left\langle \mathbf{a}_{k,t} - \left\langle \hat{\nu}_{k,t,\ell}^{-1} \hat{\mathbb{E}} \left[\phi_k \left(\left\{ \frac{\hat{s}_{k,t,\ell}}{\hat{\sigma}_{k,t,\ell}} \right\}_k \right) \frac{\mathbf{x}_{k,t,\ell}}{\hat{\sigma}_{k,t,\ell}} \right] \right\rangle_\ell \right\rangle_t, \quad (38)$$

$$\mathbf{a}_{k,t} = \frac{\left\langle \hat{\mathbf{C}}_{k,t,\ell} \right\rangle_\ell \mathbf{w}_k}{\mathbf{w}_k^H \left\langle \hat{\mathbf{C}}_{k,t,\ell} \right\rangle_\ell \mathbf{w}_k}, \quad (39)$$

and

$$\mathbf{H}_k = \left\langle \left\langle \frac{\hat{\mathbf{C}}_{k,t,\ell}}{\hat{\sigma}_{k,t,\ell}^2} \right\rangle_\ell - \left\langle \frac{\hat{\rho}_{k,t,\ell} \hat{\mathbf{C}}_{k,t,\ell}}{\hat{\nu}_{k,t,\ell}^* \hat{\sigma}_{k,t,\ell}^2} \right\rangle_\ell \right\rangle_t, \quad (40)$$

$$\mathbf{H}_k = - \left\langle \left\langle \frac{\hat{\rho}_{k,t,\ell} \hat{\mathbf{C}}_{k,t,\ell}}{\hat{\nu}_{k,t,\ell}^* \hat{\sigma}_{k,t,\ell}^2} \right\rangle_\ell \right\rangle_t, \quad (41)$$

for FastDIVA and QuickIVE, respectively.

F. Relation to Previous Methods

For $L = 1$, the update rule (37) with (38) and (40) is readily simplified to 45 in [14]. It means that FastDIVA proposed in this paper is an extension of the previous method for $L > 1$. It also follows that the proposed algorithm is the successor of FastICA from [12] (only $L = T = K = 1$) and of FastIVA from [57] (only $L = T = 1$).

Similarly, the proposed QuickIVE is the extension of the method from [56] for $L > 1$. QuickIVE provides an alternative to FastDIVA. It is an algorithm whose convergence is slightly slower than that of FastDIVA; nevertheless, this algorithm sometimes appears to be more stable, as will be shown in Section V; see also [56] where QuickIVE is shown to take an advantage over FastDIVA in continuous on-line source extraction.

⁵By neglecting the mixed second-order derivatives of (7), the developed algorithms are substantially simplified because the update rules become decoupled in k .

IV. EXTENSIONS

The current section is devoted to the Gaussian SOI source model, which comes into play when $L > 1$ (unlike for $L = 1$, the Gaussian SOI can be identified when $L > 1$). The Gaussian source model brings two important advantages. First, its analytic form leads to simplifications of both mathematical expressions and algorithms, which then operate purely with the second-order statistics: covariance (and pseudo-covariance) matrices. This is useful for capturing non-circularity and dependencies among the components of SOI when $K > 1$. Second, in situations with a critical lack of data due to very short sub-blocks, e.g., when $K \gg N_s$, an efficient source model can be designed based on an appropriate model of the SOI covariance matrix.

A. Gaussian Score Function

Let the distribution of $\mathbf{s}_{t,\ell}$ be Gaussian with covariance matrix $\Sigma_{t,\ell} = \mathbb{E}[\mathbf{s}_{t,\ell}\mathbf{s}_{t,\ell}^H]$ and pseudo-covariance matrix $\Gamma_{t,\ell} = \mathbb{E}[\mathbf{s}_{t,\ell}\mathbf{s}_{t,\ell}^T]$. From now on, we omit the indices k and t to simplify our notation, keeping in mind that the signals and their parameters are always block- and sub-block-dependent.

The log-density of the Gaussian SOI, represented by the vector \mathbf{s} , can be written in the form

$$\log f(\mathbf{s}|\Sigma, \Gamma) = -\mathbf{s}^H \mathbf{P}^{-*} \mathbf{s} + \Re \{ \mathbf{s}^T \mathbf{M}^T \mathbf{P}^{-*} \mathbf{s} \} + \text{const.}, \quad (42)$$

where $\mathbf{P} = \Sigma^* - \Gamma^H \Sigma^{-1} \Gamma$, and $\mathbf{M} = \Gamma^H \Sigma^{-1}$ [58]; \mathbf{P}^{-*} is a short notation for matrix inverse and conjugate value. Note that since $\Sigma = \Sigma^H$ and $\Gamma = \Gamma^T$, it holds that $\mathbf{P} = \mathbf{P}^H$.

Let $\psi(\mathbf{s}|\Sigma, \Gamma)$ denote the vector score function of \mathbf{s} , whose k th element is the k th score function of \mathbf{s} , i.e., $\psi_k(\mathbf{s}) = \mathbf{e}_k^H \psi(\mathbf{s}|\Sigma, \Gamma)$; \mathbf{e}_k is the k th column of \mathbf{I}_K . By definition, it holds that

$$\begin{aligned} \psi(\mathbf{s}|\Sigma, \Gamma) &= -\frac{\partial \log f(\mathbf{s})}{\partial \mathbf{s}} \\ &= \mathbf{P}^{-1} \mathbf{s}^* - \frac{1}{2} (\mathbf{M}^T \mathbf{P}^{-*} + \mathbf{P}^{-1} \mathbf{M}) \mathbf{s}. \end{aligned} \quad (43)$$

Let us consider the following well-known special cases.

- For the circular case, $\Gamma = \mathbf{M} = \mathbf{0}$ holds and the vector score function takes on a simple form $\psi(\mathbf{s}) = (\Sigma^{-1} \mathbf{s})^*$.
- Let us consider the scalar case $K = 1$, $\Sigma = \sigma^2 = 1$, and let us denote $\delta = \Gamma$. It then holds that $|\delta| \leq 1$. The score function takes on the form

$$\psi(s) = \frac{1}{1 - |\delta|^2} (s^* - \delta^* s). \quad (44)$$

The following Lemma will be useful for incorporating the Gaussian source model into the algorithms presented in the previous section.

Lemma 5: Let \mathbf{s} be the K -dimensional Gaussian vector random variable with zero mean, covariance Σ , pseudo-covariance Γ , and score function $\psi(\mathbf{s}|\Sigma, \Gamma)$. By the transformation theorem, $\phi(\mathbf{s}) = \psi(\mathbf{s}|\Lambda \Sigma \Lambda, \Lambda \Gamma \Lambda)$ is the score function of the normalized variable $\Lambda \mathbf{s}$ where $\Lambda = \text{diag}[\sigma_1^{-1}, \dots, \sigma_K^{-1}]$; $\text{diag}(\cdot)$ denotes the diagonal matrix with the values of the argument on its main diagonal. By definitions of (33) and (34), it holds that,

for $k = 1, \dots, K$,

$$\nu_k = 1, \quad (45)$$

$$\rho_k = \sigma_k^2 (\mathbf{P}^{-1})_{kk}. \quad (46)$$

Next, when Σ, Γ , and Λ are estimated, respectively, by $\tilde{\Sigma}, \tilde{\Gamma}$, and $\tilde{\Lambda}$, $\phi(\mathbf{s}) = \psi(\mathbf{s}|\tilde{\Lambda} \tilde{\Sigma} \tilde{\Lambda}, \tilde{\Lambda} \tilde{\Gamma} \tilde{\Lambda})$ is then used as the model score function for the available samples of \mathbf{s} and for $k = 1, \dots, K$,

$$\hat{\nu}_k = 1, \quad (47)$$

$$\hat{\rho}_k = \hat{\sigma}_k^2 \left(\tilde{\mathbf{P}}^{-1} \right)_{kk}, \quad (48)$$

where $\tilde{\mathbf{P}} = \tilde{\Sigma}^* - \tilde{\Gamma}^H \tilde{\Sigma}^{-1} \tilde{\Gamma}$.

Proof: See Appendix C. ■

FastDIVA and QuickIVE based on the Gaussian source model are obtained when (43), (47), and (48) are used in (38), (40), and (41). The following three subsections consider particular variants of these algorithms.

B. Scalar Gaussian SOI

Here, we consider the special case corresponding to the fundamental static BSE problem when $K = T = 1$ with the nonstationary Gaussian source model, i.e., $L > 1$. We discuss the properties of stationary points of the contrast function, the simplified learning rules of FastDIVA and QuickIVE, and compare the circular and non-circular cases.

Let us consider the circular case first, where the score function is $\psi(s) = s^*$; it follows from (44) when $\delta = 0$. The gradient (18) is then obtained in the form

$$\nabla = \mathbf{a} - \left\langle \frac{\hat{\mathbf{E}}[\hat{s}_\ell^* \mathbf{x}_\ell]}{\hat{\sigma}_\ell^2} \right\rangle_\ell = \mathbf{a} - \langle \mathbf{a}_\ell \rangle_\ell, \quad (49)$$

where we used the definitions of (11) and (15). By putting (49) equal to zero, and using that $\hat{\sigma}_\ell^2 = \mathbf{w}^H \hat{\mathbf{C}}_\ell \mathbf{w}$, we obtain an elegant form of the condition for the stationary point of the contrast function (8)

$$\frac{\langle \hat{\mathbf{C}}_\ell \rangle_\ell \mathbf{w}}{\langle \mathbf{w}^H \hat{\mathbf{C}}_\ell \mathbf{w} \rangle_\ell} = \left\langle \frac{\hat{\mathbf{C}}_\ell \mathbf{w}}{\mathbf{w}^H \hat{\mathbf{C}}_\ell \mathbf{w}} \right\rangle_\ell. \quad (50)$$

By considering $N \rightarrow +\infty$, it is seen that the SOI cannot be extracted when its variance σ_ℓ^2 is constant over ℓ , because any \mathbf{w} satisfies this condition. This observation is in agreement with the identifiability condition of the corresponding BSE problem [18], [31].

In the circular case, the learning rule of FastDIVA resp. QuickIVE is simplified to

$$\mathbf{w}^{\text{new}} = \mathbf{w} - \mathbf{H}^{-1} (\mathbf{a} - \langle \mathbf{a}_\ell \rangle_\ell), \quad (51)$$

where

$$\mathbf{H} = \frac{\langle \hat{\mathbf{C}}_\ell \rangle_\ell}{\langle \hat{\sigma}_\ell^2 \rangle_\ell} - \left\langle \frac{\hat{\mathbf{C}}_\ell}{\hat{\sigma}_\ell^2} \right\rangle_\ell \quad \text{resp.} \quad \mathbf{H} = - \left\langle \frac{\hat{\mathbf{C}}_\ell}{\hat{\sigma}_\ell^2} \right\rangle_\ell. \quad (52)$$

It is seen that, when $\hat{\sigma}_\ell^2$ tends to be constant over ℓ , the Hessian matrix of FastDIVA is close to zero, so the algorithm will have

unstable behavior in the vicinity of the SOI. By taking into account non-circularity, the above-mentioned learning rule is changed to

$$\mathbf{w}^{\text{new}} = \mathbf{w} - \mathbf{H}^{-1} \left(\mathbf{a} - \left\langle \frac{1}{1 - |\hat{\delta}_\ell|^2} \left(\mathbf{a}_\ell - \hat{\delta}_\ell^* \frac{\hat{\mathbf{D}}_\ell \mathbf{w}^*}{\hat{\sigma}_\ell^2} \right) \right\rangle_\ell \right), \quad (53)$$

where $\hat{\mathbf{D}}_\ell = \hat{\mathbf{E}}[\mathbf{x}_\ell \mathbf{x}_\ell^T]$ is the sample pseudo-covariance matrix of \mathbf{x}_ℓ , and $\hat{\delta}_\ell = \mathbf{w}^H \hat{\mathbf{D}}_\ell \mathbf{w}^* / \hat{\sigma}_\ell^2$ is the normalized circularity coefficient of the SOI satisfying $|\hat{\delta}_\ell| \leq 1$, and

$$\mathbf{H} = \frac{\langle \hat{\mathbf{C}}_\ell \rangle_\ell}{\langle \hat{\sigma}_\ell^2 \rangle_\ell} - \left\langle \frac{\hat{\mathbf{C}}_\ell}{(1 - |\hat{\delta}_\ell|^2) \hat{\sigma}_\ell^2} \right\rangle_\ell, \quad (54)$$

and

$$\mathbf{H} = - \left\langle \frac{\hat{\mathbf{C}}_\ell}{(1 - |\hat{\delta}_\ell|^2) \hat{\sigma}_\ell^2} \right\rangle_\ell, \quad (55)$$

for FastDIVA and QuickIVE, respectively.

We can see that, for $\hat{\delta}_\ell = 0$, $\ell = 1, \dots, L$, the algorithms given by (51) and (53) coincide. However, in experiments, we have observed that the latter update rule appears to be numerically more stable than the former one, because $\hat{\delta}_\ell$ s are never exactly equal to zero even when the SOI is circular.

The above-described algorithms can be easily extended to the nonstationary mixing conditions when $T > 1$ by replacing the gradient and Hessian matrix by their averages over the blocks, as follows from the general formulas (38) and (40).

C. Vector Gaussian SOI: General Covariance Structure

In the case of $K \geq 1$, the Gaussian source model can be directly applied to derive the update rules of FastDIVA and QuickIVE by considering the Gaussian model score function given by (43). This approach is reasonable when no prior knowledge about $\mathbf{\Sigma}$ or $\mathbf{\Gamma}$ is available. The sample-based estimates $\tilde{\mathbf{\Sigma}} = \hat{\mathbf{E}}[\hat{\mathbf{s}}\hat{\mathbf{s}}^H]$ and $\tilde{\mathbf{\Gamma}} = \hat{\mathbf{E}}[\hat{\mathbf{s}}\hat{\mathbf{s}}^T]$ can be used where \mathbf{s} stands for the current estimate of the SOI. We can then apply (47) and (48) in Lemma 5 with $\tilde{\mathbf{\Sigma}} = \tilde{\mathbf{\Sigma}}$ and $\tilde{\mathbf{\Gamma}} = \tilde{\mathbf{\Gamma}}$; subsequently we can replace the unknown matrices in (48) by their estimates, and put them into (38), (40), and (41).

However, two practical issues occur when K gets “larger”. First, this approach tends to be stable and accurate until a critical number of samples within sub-blocks is available. The problem arises when $K \approx N_s$ because of the rank deficiency of $\tilde{\mathbf{\Sigma}}$. Obviously, $\tilde{\mathbf{\Sigma}}$ is singular when $K > N_s$. It can be avoided by adding a regularizing term to $\tilde{\mathbf{\Sigma}}$ such as a multiple of the identity matrix, that is,

$$\tilde{\mathbf{\Sigma}} = \hat{\mathbf{E}}[\hat{\mathbf{s}}\hat{\mathbf{s}}^H] + \mu \mathbf{I}_K. \quad (56)$$

Here, the parameter $\mu \geq 0$ provides a trade-off between the source modeling accuracy and algorithm stability.

The second issue is that the computational complexity steeply grows with K due to the computations of $\tilde{\mathbf{\Sigma}}^{-1}$ and $\tilde{\mathbf{\Gamma}}^{-1}$. Although these matrices are Hermitian and positive definite, the computations still take at least $\mathcal{O}(K^3)$ operations. This brings about a prohibitively large computational burden in some

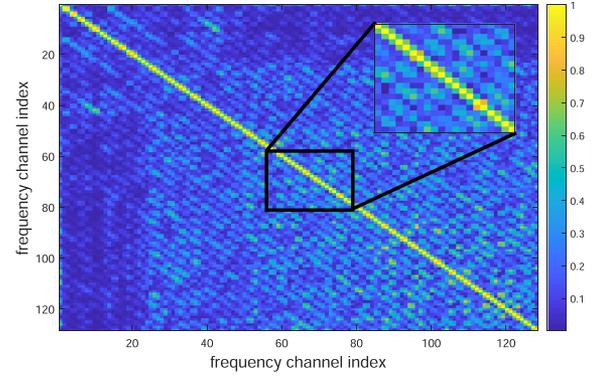


Fig. 1. Example of a typical covariance matrix of normalized STFT channels of speech; a clear male speech sampled at 16 kHz; FFT length of 256 samples; window shift of 128 samples; the Hamming analysis window used; average taken over 50 frames.

applications, such as in the frequency-domain audio source separation where K corresponds to the frequency resolution (e.g., $K \geq 128$). Moreover, no significant algebraic simplifications in (38) are possible unless these matrices have any favourable structure. Altogether, the approach discussed in this subsection is recommended only for “small” enough values of K .

D. Vector Gaussian SOI: Tridiagonal Covariance Matrix

There are situations when $\tilde{\mathbf{\Sigma}}$ and $\tilde{\mathbf{\Gamma}}$ are structured, which can be used to alleviate the shortcomings of the previous approach. In this subsection, we consider the special case in which $\tilde{\mathbf{\Sigma}}$ is tridiagonal and $\tilde{\mathbf{\Gamma}} = \mathbf{0}$ (non-circularity is not taken into account). Without any loss of generality, we will consider $\tilde{\mathbf{\Sigma}}$ when all SOI components are normalized to have unit sample variance, so the assumed structure is given by

$$\tilde{\mathbf{\Sigma}} = \begin{pmatrix} 1 & f_1 & & & \\ f_1^* & 1 & f_2 & & \\ & f_2^* & \ddots & \ddots & \\ & & \ddots & \ddots & f_{K-1} \\ & & & f_{K-1}^* & 1 \end{pmatrix}. \quad (57)$$

This structure means that adjacent components of the SOI are correlated. It is motivated by the situation that appears, e.g., in speech extraction in the short-term Frequency domain (STFT). Fig. 1 shows a typical sample covariance matrix of normalized STFT channels of speech. There are significant correlations of adjacent frequency bands caused by their overlap; the other correlations appear to be less significant.

By taking the advantage of this structure, the close-form formula from [59] can now be used to compute $\tilde{\mathbf{\Sigma}}^{-1}$, which reduces the main computational burden needed for the evaluation of (48) and (43). We have that

$$(\tilde{\mathbf{\Sigma}}^{-1})_{ij} = \begin{cases} (-1)^{i+j} f_i \dots f_{j-1} \theta_{i-1} \xi_{j+1} / \theta_K & i < j \\ \theta_{i-1} \xi_{j+1} / \theta_K & i = j \\ (-1)^{i+j} f_j^* \dots f_{i-1}^* \theta_{j-1} \xi_{i+1} / \theta_K & j < i \end{cases} \quad (58)$$

for $i, j = 1, \dots, K$, and

$$\theta_i = \theta_{i-1} - |f_{i-1}|^2 \theta_{i-2}, \quad i = 2, 3, \dots, K, \quad (59)$$

with initial conditions $\theta_0 = \theta_1 = 1$, and

$$\xi_i = \xi_{i+1} - |f_i|^2 \xi_{i+2}, \quad i = K-1, \dots, 1, \quad (60)$$

with initial conditions $\xi_{K+1} = \xi_K = 1$. Using this, the computational burden due to $\tilde{\Sigma}^{-1}$ is substantially reduced to almost linear complexity as follows.

Since $|f_i| < 1$, (58) means that the off-diagonal entries of $\tilde{\Sigma}^{-1}$ are exponentially decreasing with the growing distance from the main diagonal. We can therefore neglect the elements of $\tilde{\Sigma}^{-1}$ on the k th diagonal for $|k| > k_{\max}$. The evaluation of $\tilde{\Sigma}^{-1}$ then only takes $\mathcal{O}(k_{\max}K)$ operations. Also, the multiplication by $\tilde{\mathbf{P}}^{-1}$ in (43), which otherwise costs $\mathcal{O}(K^2)$, is reduced to $\mathcal{O}(k_{\max}K)$.

Another issue is that the positive definiteness of $\tilde{\Sigma}$ must be ensured for stability of the algorithms. Here, we propose to constrain the off-diagonal entries of $\tilde{\Sigma}$ as

$$f_k = \begin{cases} \hat{f}_k & |\hat{f}_k| \leq 0.4 \\ 0.4 \cdot \frac{\hat{f}_k}{|\hat{f}_k|} & |\hat{f}_k| > 0.4 \end{cases}, \quad (61)$$

where $\hat{f}_k = \hat{\mathbf{E}}[\hat{s}_k \hat{s}_{k+1}^*]$; $k = 1, \dots, K-1$. The threshold for limiting the magnitude of \hat{f}_k by 0.4 is inspired by the analytic value of the eigenvalues of tridiagonal matrices where f_k s are all constant and equal to f . Their eigenvalues are

$$1 + 2|f| \cos\left(\frac{k\pi}{K+1}\right), \quad k = 1, \dots, K. \quad (62)$$

In that case, the limit 0.4 hence ensures that all eigenvalues of $\tilde{\Sigma}$ are sufficiently larger than zero.

V. EXPERIMENTAL VALIDATION

A. Simulations

The proposed algorithms are validated in simulated experiments and compared with other state-of-the-art algorithms. Their performance is assessed in terms of the interference-to-signal ratio (ISR) measured on the extracted signal(s). The 1% trimmed mean is used for averaging over Monte Carlo repetitions 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.

In a simulation trial, the SOI(s) samples are drawn independently according to the complex Generalized Gaussian law [60] with zero mean, (normalized) circularity coefficient $|\delta| \leq 1$, and the shape parameter $c > 0$ ($c = 1$ means Gaussian, $c < 1$ super-Gaussian, and $c > 1$ sub-Gaussian). The nonstationarity of the SOI is driven through its variance, which is, on the t th block and ℓ th subblock, equal to

$$\sigma_{t,\ell}^2 = \left[\sin\left(\frac{t\pi}{T+1}\right) \sin\left(\frac{\ell\pi}{L+1}\right) \right]^\alpha. \quad (63)$$

The SOI is stationary for $\alpha = 0$, moderately dynamical for $\alpha \approx 1$, and transient-like when $\alpha \gg 1$; see the example shown

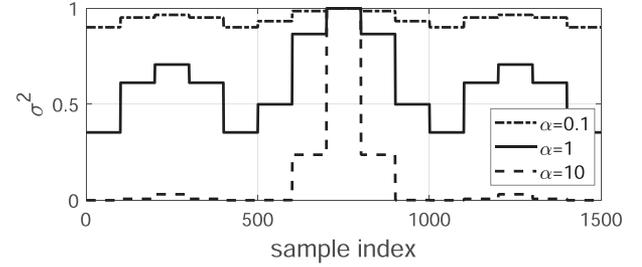


Fig. 2. Example of variance profiles of the SOI according to (63) when $T = 3$, $L = 5$, $N_s = 100$.

in Fig. 2. The background sources are generated as stationary uncorrelated circular Gaussian.

The signals are mixed, in each block, by a random mixing matrix. The mixing matrices are generated as suggested in [13], i.e., their inverse matrices are generated from the complex-valued uniform distribution on $[1, 2]$ (real and imaginary values); for $T > 1$, the CSV mixing model is ensured by making the first rows of the inverse matrices constant over t . The initial signal-to-interference ratio on input channels is chosen to be approximately constant [13]. The algorithms are initialized by the separating vectors $\mathbf{w}_k^{\text{ini}} = \mathbf{w}_k^* + \boldsymbol{\epsilon}_k$ where $\boldsymbol{\epsilon}_k$ is a random vector orthogonal to \mathbf{w}_k^* such that $\|\boldsymbol{\epsilon}_k\|^2 = 0.01$.

1) *Static Independent Component Extraction*: The standard ICE problem with $T = K = 1$ is considered here where $L = 20$, $d = 6$, and $N = 5000$, i.e., $N_s = 250$. The SOI distribution is generated with $c = 1$ and $\delta = 0.5$ (non-circular Gaussian); the case when $c = 0.5$ (non-circular Laplacean) is provided in the supplementary material of this article.

We have compared eight algorithms. BOGIVE_w [48], [61], FastDIVA from [14], and CSV-AuxIVE [48] represent non-Gaussianity-based methods that inherently work with the hypothesis that $L = 1$. In BOGIVE_w and FastDIVA, the rational nonlinearity $\phi(s) = \frac{s^*}{1+|s|^2}$ denoted as “rati” is used; CSV-AuxIVE utilizes the standard nonlinearity for super-Gaussian sources [25]. As for a method assuming Gaussianity and non-circularity and allowing for nonstationarity, we compare LLJBD from [20]. The extended versions of FastDIVA and QuickIVE proposed in this article stand for the methods allowing for the non-Gaussianity and/or nonstationarity. They are tested with $L = 20$ and with the nonlinearities “rati” or “gauss,” where the latter corresponds to (44), i.e., the Gaussian score allowing for non-circularity. From here, short abbreviations *algorithm–nonlinearity–L* will be used for the variants of FastDIVA and QuickIVE, e.g., FastDIVA–rati–20.

Fig. 3 shows the results when $c = 1$ for 1,000 trials. For $\alpha = 0.1$, the algorithms yield poor ISR (> -15 dB) since the SOI is almost stationary and Gaussian, up to FastDIVA–gauss–20 and QuickIVE–gauss–20 that benefit from the non-circularity of the SOI. With growing α , all algorithms are improving as the SOI becomes more nonstationary, including the non-Gaussianity-based FastDIVA–rati–1 and BOGIVE_w. However, BOGIVE_w fails to achieve the same performance as FastDIVA due to its slow convergence (it is limited by the maximum number of 1,000 iterations).

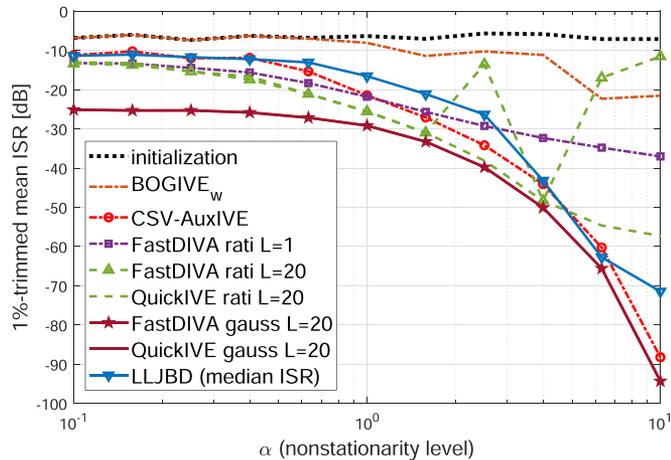


Fig. 3. Resulting ISR as a function of α , the parameter controlling the nonstationarity of the SOI according to (63). The pdf of the SOI is Gaussian as $c = 1$ with circularity coefficient $\delta = 0.5$; “initialization” corresponds to the do-nothing approach and reflects the ISR given by the initialization.

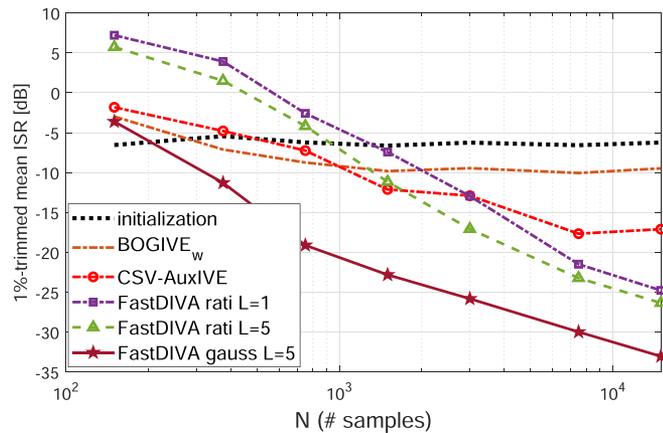


Fig. 4. ISR as a function of N when $c = 1$ and $\delta = 0.5$ (Gaussian non-circular SOI), $T = 3$ (nonstationary CSV mixing model), and $L = 5$ and $\alpha = 2$ (nonstationary source model). N ranges from 150 through 15,000; the length of sub-blocks N_s thus ranges from 15 through 1,000.

The performance is better for algorithms capturing the nonstationarity considering $L = 20$; especially, for $\alpha > 1$. The superior performance is achieved by FastDIVA–gauss–20 and QuickIVE–gauss–20, whose performance characteristics coincide in this case. The superiority is achieved due to the accurate source modelling including non-circularity. These methods yield lower performance with the “rati” nonlinearity (assuming circularity); FastDIVA–rati–20 shows less stable convergence than QuickIVE–rati–20 for the values of $\alpha > 1$.

LLJBD is also improving with growing α ; however, the median ISR has to be shown here due to unstable convergence. Surprising results are obtained by CSV-AuxIVE since it shows improvement with growing α similarly to the methods employing nonstationarity. This is in contrast with the fact that the method comes from the optimization of non-Gaussianity-based source model [25], [48]. The theoretical explanation of this behavior goes beyond the scope this paper.

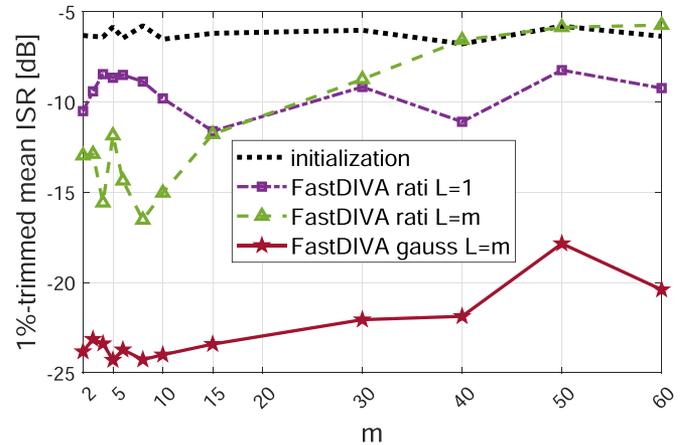


Fig. 5. ISR as a function of m , where m is the number of sub-blocks assumed within the proposed algorithms; $c = 1$, $\delta = 0.5$ (Gaussian non-circular SOI), $T = 3$ (nonstationary CSV mixing model), $L = 5$, $\alpha = 2$ (nonstationary source model), $N = 1,800$; m ranges from 2 through 60; N_s thus ranges from 300 through 10.

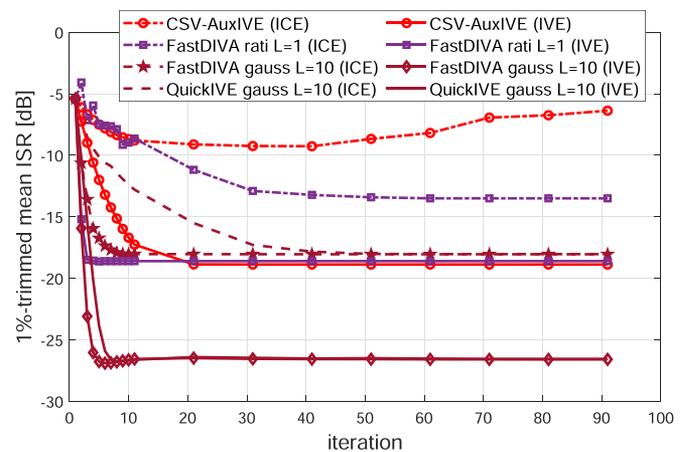


Fig. 6. Mean convergence of algorithms performing separate (ICE) and joint (IVE) blind extraction: ISR as a function of iteration index. The parameters of the experiment are $\alpha = 2$, $c = 0.5$, $\delta = 0.5$, $L = 10$ (nonstationary Laplacean non-circular SOI), $d = 10$, $T = 1$ (static mixtures), $K = 5$ (five jointly dependent mixtures per trial), and $N = 500$.

2) *Dynamic Independent Component Extraction*: We now turn to $T = 3$, $K = 1$, $d = 6$, and where the SOI is nonstationary and non-circular Gaussian with $L = 5$, $\alpha = 2$ and $\delta = 0.5$. Since $T > 1$, we compare only the methods that allow for the CSV mixing model. QuickIVE is not presented here since it has provided the same results as FastDIVA.

Fig. 4 shows the average ISR achieved after 1,000 trials as a function of the length of data N . The values of N are selected so that N_s ranges from the extremely small value of $N_s = 15$ through $N_s = 1,000$. With growing N , all methods are improving, including BOGIVE_w, CSV-AuxIVE, and FastDIVA–rati–1, which do not exploit the nonstationarity of the SOI on the sub-blocks. The extended FastDIVA (allowing for $L = 5$) shows better performance than with $L = 1$. Moreover, FastDIVA–gauss–20 takes the advantage of non-circularity and

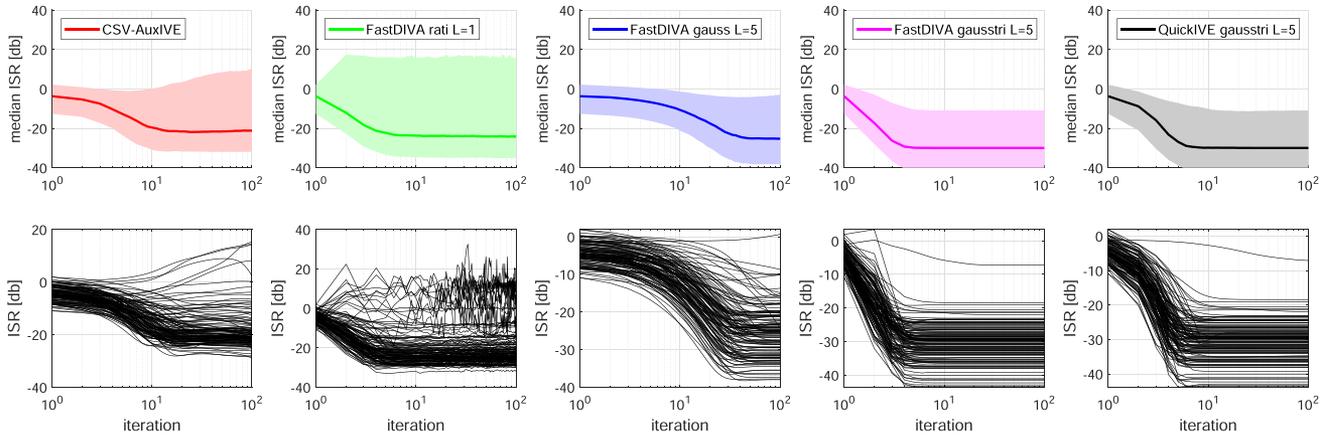


Fig. 7. Results of the simulated frequency-domain speech extraction. Row 1: median ISR taken over 100 trials; the average ISR (line); the range from the minimum through the maximum value of ISR over all frequencies (area). Row 2: the ISR of all frequencies in trial 1.

achieves a useful accuracy level (of ≈ -11 dB) even in the extreme case of $N_s = 25$ (resp. $N = 150$). The experiment when the SOI is Laplacean is shown in the supplementary material.

Next, we examine the sensitivity of the algorithms to the assumed value of L . The setup is the same as in the previous experiment with $N = 1,800$. The true number of sub-blocks is $L = 5$ while the proposed algorithms are run with $L = m$ where m ranges between 2 and 60; the number of samples within sub-blocks N_s thus ranges from 300 through 10. Fig. 5 compares three variants of FastDIVA. FastDIVA-rati-1 is independent of m , so its performance only fluctuates between -8 and -12 dB of mean ISR. FastDIVA-rati- m shows quite stable performance until $m \leq 15$ ($N_s \geq 40$), and its performance is deteriorating for $m > 15$. FastDIVA-gauss- m shows a stable performance; it only slightly deteriorates when $m \geq 50$ (critically low $N_s \leq 12$). In conclusion, the results show that the algorithms are robust enough in the model parameter L until N_s is sufficiently high.

3) *Independent Vector Extraction*: Here, we consider $K = 5$ mixtures of dimension $d = 10$ involving jointly dependent components of the SOI. These components are generated as follows: First, five signals are generated independently with $\alpha = 2$, $c = 0.5$, $\delta = 0.5$, $L = 10$ (nonstationary Laplacean non-circular sources). Second, these signals are multiplied by a random $K \times K$ matrix drawn from $\mathcal{CN}(0, 1)$, which yields dependent and correlated SOI components. K mixtures with $T = 1$ are generated, so the data obeys the standard IVE mixing model. The number of samples is $N = 500$; N_s is 50. The compared algorithms are tested in two regimes: the K mixtures are processed (ICE) separately and (IVE) jointly.

Fig. 6 shows ISR averaged over the mixtures and 1,000 trials as a function of iteration index. In the ICE regime, these methods show significantly slower convergence than in the IVE regime (e.g., CSV-AuxIVE or FastDIVA-rati-1) and also lower accuracy levels because the dependencies among the SOI components are not used. This shows that proper source modeling influences not only the algorithms' accuracy but also their convergence. FastDIVA is shown to be mostly the fastest method among the compared ones; QuickIVE provides its slightly decelerated (and sometimes more stable) variant.

B. Frequency-Domain Blind Speech Extraction

The typical application of IVE includes speech extraction in the short time frequency domain. The mixture of speech and background are convolutive in the time domain and can be approximated as instantaneous in the frequency domain; hence the instantaneous complex-valued ICE and IVE mixing models can be applied. Compared to ICE, IVE tries to secure that the SOI is extracted in each frequency band (the permutation problem) by using dependencies [24]. However, it is generally known that this solution through IVE is not definite. For example, the extracted frequency components can form groups corresponding to different independent sources, so, finally, the complete extraction is not achieved. The experiment proposed here aims at a detailed investigation of the convergence issues of the compared algorithms in this application. Also, the variants of FastDIVA/QuickIVE for Gaussian SOI with tridiagonal covariance matrix proposed in Section IV-D are employed here (denoted with the “gausstri” nonlinearity).

In a trial, a short interval of clean speech is transformed by the Short-Time Fourier Transform (STFT) with the FFT/window and shift length, respectively, equal to $2K + 1$ and K samples; we consider $K = 128$; the number of the STFT frames corresponds to $N = 375$. Then, K mixtures, one per each frequency band $1, \dots, K$ (the zero frequency band is not included), obeying the CSV model with $d = 10$ (simulating 10 microphones) and $T = 3$ (a moving speaker) are generated. The speech frequency components play the role of the SOI components; the background is generated from $d - 1$ independent Laplacean signals.

The results of five algorithms are evaluated in Fig. 7, where the ISR is shown as a function of iteration index. The graphs in the first row show median ISR taken over 100 trials; the lines show the average ISR taken over all K frequencies while the transparent areas show the range from the minimum through the maximum ISR over the frequencies. For more insight, the charts in the second row illustrate the results of trial 1.

The example shows that all algorithms tend to extract most of the frequency components of the speech. This is indicated by

the average median ISR whose value goes significantly below 0 dB. However, the ISR of some frequencies is sometimes growing, which seems to happen more often for CSV-AuxIVE and FastDIVA-rati-1. FastDIVA-gauss-5 (Section IV-C) appears to yield a more stable convergence in all frequencies; however, it is significantly slower than the other methods and is computationally very expensive due to $K = 128$. The fastest and most reliable convergence is observed in FastDIVA-gauss-5 and QuickIVE-gauss-5. The average median ISR by these methods achieves values below -20 dB after less than 10 iterations, which is the superior extraction accuracy among the compared methods.

VI. CONCLUSION

The BSE model, combining non-Gaussianity and nonstationarity in the source model and nonstationarity in the mixing models, makes the set of identifiable sources broader. We have derived extended variants of FastDIVA and QuickIVE, which show faster convergence and higher accuracy than the state-of-the-art methods. We have shown that these algorithms can be used with the Gaussian score function, which makes them purely based on second-order statistics. In complex-valued problems, they can efficiently exploit non-circularity. The special variant, assuming jointly Gaussian SOIs with tridiagonal covariance matrix, shows promising results for frequency-domain blind speaker extraction.

APPENDIX A PROOF OF LEMMA 1

The gradient of the first two terms in (8), here denoted as ∇_{12} , readily gives that

$$\begin{aligned} \nabla_{12} &= \frac{\partial}{\partial \mathbf{w}^H} \left\langle \hat{\mathbf{E}} \left[\log f \left(\frac{\hat{s}_\ell}{\hat{\sigma}_\ell} \right) \right] - \log \hat{\sigma}_\ell^2 \right\rangle_\ell = \\ &= \left\langle \hat{\mathbf{E}} \left[\phi \left(\frac{\hat{s}_\ell}{\hat{\sigma}_\ell} \right) \frac{\mathbf{x}_\ell}{\hat{\sigma}_\ell} \right] + \Re(\hat{\nu}_\ell) \mathbf{a}_\ell - \mathbf{a}_\ell \right\rangle_\ell, \end{aligned} \quad (64)$$

where \mathbf{a}_ℓ is defined by (15), and where we used identities

$$\frac{\partial}{\partial \mathbf{w}^H} \hat{s}_\ell = \frac{\partial}{\partial \mathbf{w}^H} \mathbf{w}^H \mathbf{x}_\ell = \mathbf{x}_\ell, \quad (65)$$

$$\frac{\partial}{\partial \mathbf{w}^H} \frac{1}{\hat{\sigma}_\ell} = \frac{\partial}{\partial \mathbf{w}^H} \frac{1}{\sqrt{\mathbf{w}^H \hat{\mathbf{C}}_\ell \mathbf{w}}} = -\frac{\mathbf{a}_\ell}{2\hat{\sigma}_\ell}, \quad (66)$$

$$\frac{\partial}{\partial \mathbf{w}^H} \log \hat{\sigma}_\ell^2 = \frac{\partial}{\partial \mathbf{w}^H} \log \mathbf{w}^H \hat{\mathbf{C}}_\ell \mathbf{w} = \mathbf{a}_\ell. \quad (67)$$

We continue by showing that the gradient of the last two terms in (8), denoted as ∇_{34} , reads

$$\nabla_{34} = \frac{\partial}{\partial \mathbf{w}^H} \left\langle -\hat{\mathbf{E}} \left[\hat{\mathbf{z}}_\ell^H \mathbf{R}_\ell \hat{\mathbf{z}}_\ell \right] + (d-2) \log |\gamma|^2 \right\rangle_\ell = \mathbf{a}. \quad (68)$$

To this end, note that the two terms inside the averaging operator $\langle \cdot \rangle_\ell$ in (8) have the same analytic shape as the last two terms in 26 in [13]. Therefore, we can employ the result given by 27 in [13], taking into account the linearity of the operator $\langle \cdot \rangle_\ell$ and the dependency of the signals and their statistics on

ℓ . By considering substitutions $\mathbf{z} \rightarrow \mathbf{z}_\ell$, $\hat{\mathbf{C}}_{\mathbf{z}} \rightarrow \hat{\mathbf{C}}_{\mathbf{z}}^\ell$, $\mathbf{R} \rightarrow \mathbf{R}_\ell$, we obtain that

$$\begin{aligned} \nabla_{34} &= 2\mathbf{a} \operatorname{tr} \left(\mathbf{R}_\ell \left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell \right) \\ &= \left(\mathbf{w}^H \hat{\mathbf{C}} \mathbf{w} \right)^{-1} \left(\hat{\mathbf{C}} \mathbf{E}^H \mathbf{R}_\ell \left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell \mathbf{h} - \operatorname{tr}(\mathbf{R}_\ell \hat{\mathbf{B}} \hat{\mathbf{C}} \mathbf{E}^H) \hat{\mathbf{C}} \mathbf{e}_1 \right) \\ &= 2(d-2)\mathbf{a} + (\gamma^*)^{-1}(d-2)(\mathbf{w}^H \hat{\mathbf{C}} \mathbf{w})^{-1} \hat{\mathbf{C}} \mathbf{e}_1, \end{aligned} \quad (69)$$

where $\operatorname{tr}(\cdot)$ denotes the trace; $\mathbf{E} = [\mathbf{0} \ \mathbf{I}_{d-1}]$; \mathbf{e}_1 denotes the first column of \mathbf{I}_d . Now, we put $\mathbf{R}_\ell = \langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \rangle_\ell^{-1}$ as assumed by the Lemma, (69) turns to

$$\begin{aligned} \nabla_{34} &= 2\mathbf{a} - \left(\mathbf{w}^H \hat{\mathbf{C}} \mathbf{w} \right)^{-1} \left(\hat{\mathbf{C}} \mathbf{E}^H \mathbf{h} \right. \\ &= -\operatorname{tr} \left(\left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell^{-1} \hat{\mathbf{B}} \hat{\mathbf{C}} \mathbf{E}^H \right) \hat{\mathbf{C}} \mathbf{e}_1 \left. \right) \\ &+ (\gamma^*)^{-1}(d-2) \left(\mathbf{w}^H \hat{\mathbf{C}} \mathbf{w} \right)^{-1} \hat{\mathbf{C}} \mathbf{e}_1. \end{aligned} \quad (70)$$

By similar steps to those given by 31 and 32 in [13],

$$\begin{aligned} \left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell^{-1} \hat{\mathbf{B}} \hat{\mathbf{C}} &= \left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell^{-1} \mathbf{B} \left\langle \hat{\mathbf{E}} \left[\mathbf{x}_\ell \mathbf{x}_\ell^H \right] \right\rangle_\ell \\ &= \left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell^{-1} \left\langle \hat{\mathbf{E}} \left[\mathbf{z}_\ell \left[\hat{s}_\ell^* \ \hat{\mathbf{z}}_\ell^H \right] \right] \right\rangle_\ell \mathbf{A}^H \\ &= \left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell^{-1} \left[\mathbf{0} \ \left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell \right] \mathbf{A}^H = \mathbf{E} \mathbf{A}^H, \end{aligned} \quad (71)$$

where we have used that $\hat{\mathbf{E}}[\mathbf{z}_\ell \hat{s}_\ell^*] = \mathbf{0}$, which follows from (10). Hence,

$$\operatorname{tr} \left(\left\langle \hat{\mathbf{C}}_{\mathbf{z}}^\ell \right\rangle_\ell^{-1} \hat{\mathbf{B}} \hat{\mathbf{C}} \mathbf{E}^H \right) = \operatorname{tr}(\mathbf{E} \mathbf{A}^H \mathbf{E}^H) = -\beta - (d-2)(\gamma^*)^{-1}. \quad (72)$$

By putting (72) into (71) and using (11), (69) follows. The assertion of the Lemma follows by summing (64) and (68).

APPENDIX B PROOF OF LEMMA 3 AND 4

The proof follows analogous steps to those in Appendix A in [14]. For $N \rightarrow +\infty$, the sample-based estimates are replaced by the expectation values, and, by definition,

$$\mathbf{H}_1 = \frac{\partial \nabla^H}{\partial \mathbf{w}} = \left[\frac{\partial}{\partial \mathbf{w}^H} \left(\mathbf{a}^T - \left\langle \nu_\ell^{-1} \mathbf{E} \left[\phi \left(\frac{s_\ell}{\sigma_\ell} \right) \frac{\mathbf{x}_\ell^T}{\sigma_\ell} \right] \right\rangle_\ell \right) \right]^* \quad (73)$$

$$\mathbf{H}_2 = \frac{\partial \nabla^T}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \left(\mathbf{a}^T - \left\langle \nu_\ell^{-1} \mathbf{E} \left[\phi \left(\frac{s_\ell}{\sigma_\ell} \right) \frac{\mathbf{x}_\ell^T}{\sigma_\ell} \right] \right\rangle_\ell \right). \quad (74)$$

The dependent variables on \mathbf{w} are $s_\ell = \mathbf{w}^H \mathbf{x}_\ell$ and $\sigma_\ell^2 = \mathbf{w}^H \mathbf{C}_\ell \mathbf{w}$; ν_ℓ are treated as constants. For proving both of the Lemmas, the following identities will be used:

$$\frac{\partial}{\partial \mathbf{w}} \frac{1}{\sigma_\ell} = -\frac{\mathbf{a}_\ell^*}{2\sigma_\ell}, \quad \frac{\partial}{\partial \mathbf{w}^H} \frac{1}{\sigma_\ell} = -\frac{\mathbf{a}_\ell}{2\sigma_\ell}, \quad (75)$$

$$\frac{\partial}{\partial \mathbf{w}} \frac{s_\ell^*}{\sigma_\ell} = \frac{\mathbf{x}_\ell^*}{\sigma_\ell} - \frac{s_\ell^* \mathbf{a}_\ell^*}{2\sigma_\ell}, \quad \frac{\partial}{\partial \mathbf{w}^H} \frac{s_\ell}{\sigma_\ell} = \frac{\mathbf{x}_\ell}{\sigma_\ell} - \frac{s_\ell \mathbf{a}_\ell}{2\sigma_\ell}, \quad (76)$$

and

$$\begin{aligned}\frac{\partial}{\partial \mathbf{w}^H} \phi_\ell \frac{\mathbf{x}_\ell^T}{\sigma_\ell} &= \left(\frac{\partial \phi_\ell}{\partial s_\ell} \left(\frac{\mathbf{x}_\ell}{\sigma_\ell} - \frac{s_\ell \mathbf{a}_\ell}{2\sigma_\ell} \right) - \frac{\partial \phi_\ell}{\partial s_\ell^*} \frac{s_\ell^* \mathbf{a}_\ell}{2\sigma_\ell} - \phi_\ell \frac{\mathbf{a}_\ell}{2} \right) \frac{\mathbf{x}_\ell^T}{\sigma_\ell}, \\ \frac{\partial}{\partial \mathbf{w}} \phi_\ell \frac{\mathbf{x}_\ell^T}{\sigma_\ell} &= \left(\frac{\partial \phi_\ell}{\partial s_\ell^*} \left(\frac{\mathbf{x}_\ell^*}{\sigma_\ell} - \frac{s_\ell^* \mathbf{a}_\ell^*}{2\sigma_\ell} \right) - \frac{\partial \phi_\ell}{\partial s_\ell} \frac{s_\ell \mathbf{a}_\ell^*}{2\sigma_\ell} - \phi_\ell \frac{\mathbf{a}_\ell^*}{2} \right) \frac{\mathbf{x}_\ell^T}{\sigma_\ell}.\end{aligned}$$

where ϕ_ℓ is a short notation of $\phi(\frac{s_\ell}{\sigma_\ell})$. Considering the expectation values of the latter two expressions, and the fact that $\mathbf{x}_\ell = \mathbf{a}s_\ell + \mathbf{y}_\ell$ where s_ℓ and \mathbf{y}_ℓ are independent, we obtain

$$\frac{\partial}{\partial \mathbf{w}^H} \mathbb{E} \left[\phi_\ell \frac{\mathbf{x}_\ell^T}{\sigma_\ell} \right] = \left[\eta_\ell \mathbf{a} - \frac{\tau_\ell}{2} \mathbf{a}_\ell \right] \mathbf{a}^T + \frac{\pi_\ell}{\sigma_\ell^2} \mathbf{P}_y^\ell, \quad (77)$$

$$\frac{\partial}{\partial \mathbf{w}} \mathbb{E} \left[\phi_\ell \frac{\mathbf{x}_\ell^T}{\sigma_\ell} \right] = \left[\xi_\ell \mathbf{a}^* - \frac{\tau_\ell}{2} \mathbf{a}_\ell^* \right] \mathbf{a}^T + \frac{\rho_\ell}{\sigma_\ell^2} (\mathbf{C}_y^\ell)^*, \quad (78)$$

where $\pi_\ell = \mathbb{E}[\frac{\partial \phi_\ell}{\partial s}]$, and $\mathbf{P}_y^\ell = \mathbb{E}[\mathbf{y}_\ell \mathbf{y}_\ell^T]$ and $\mathbf{C}_y^\ell = \mathbb{E}[\mathbf{y}_\ell \mathbf{y}_\ell^H]$ are the covariance and pseudo-covariance of \mathbf{y}_ℓ , respectively. Owing to the assumption stated in Section II-B that the background signals are circular Gaussian, $\mathbf{P}_y^\ell = \mathbf{0}$. Next, it holds that $\mathbf{C}_\ell = \mathbf{a}\mathbf{a}^H \sigma_\ell^2 + \mathbf{C}_y^\ell$. By making these substitutions in (77) and (78), we obtain

$$\frac{\partial}{\partial \mathbf{w}^H} \mathbb{E} \left[\phi_\ell \frac{\mathbf{x}_\ell^T}{\sigma_\ell} \right] = \left[\eta_\ell \mathbf{a} - \frac{\tau_\ell}{2} \mathbf{a}_\ell \right] \mathbf{a}^T, \quad (79)$$

$$\frac{\partial}{\partial \mathbf{w}} \mathbb{E} \left[\phi_\ell \frac{\mathbf{x}_\ell^T}{\sigma_\ell} \right] = \left[(\xi_\ell - \rho_\ell) \mathbf{a}^* - \frac{\tau_\ell}{2} \mathbf{a}_\ell^* \right] \mathbf{a}^T + \frac{\rho_\ell}{\sigma_\ell^2} \mathbf{C}_\ell^*. \quad (80)$$

By inserting (79) and (80), respectively, into (73) and (74), the assertion of Lemma 4 follows.

What is left to compute for the proof of Lemma 3 are the derivatives of \mathbf{a}^T in (73) and (74) when the OGC (11) is imposed:

$$\frac{\partial \mathbf{a}^T}{\partial \mathbf{w}^H} = -\mathbf{a}\mathbf{a}^T, \quad (81)$$

$$\frac{\partial \mathbf{a}^T}{\partial \mathbf{w}} = \frac{\mathbf{C}^*}{\sigma^2} - \mathbf{a}^* \mathbf{a}^T, \quad (82)$$

where $\sigma^2 = \langle \sigma_\ell^2 \rangle_\ell$. By inserting (79) and (81) into (73), and (80) and (81) into (74), the assertion of Lemma 3 follows.

APPENDIX C

PROOF OF LEMMA 5

By definition, $\mathbf{P} = \Sigma^* - \Gamma^H \Sigma^{-1} \Gamma$ and $\mathbf{M} = \Gamma^H \Sigma^{-1}$. Using the matrix inverse lemma, it can be shown that

$$\mathbf{P}^{-1} \Sigma^* - \frac{1}{2} (\mathbf{M}^T \mathbf{P}^{-*} + \mathbf{P}^{-1} \mathbf{M}) \Gamma = \mathbf{I}_K. \quad (83)$$

We will use this identity for proving (45).

By considering the substitutions $\Sigma \leftarrow \Lambda \Sigma \Lambda$ and $\Gamma \leftarrow \Lambda \Gamma \Lambda$ and (43), the score function of the normalized \mathbf{s} reads

$$\begin{aligned}\phi(\mathbf{s}) &= \psi(\mathbf{s} | \Lambda \Sigma \Lambda, \Lambda \Gamma \Lambda) = \Lambda^{-1} \mathbf{P}^{-1} \Lambda^{-1} \mathbf{s}^* \\ &\quad - \frac{1}{2} \Lambda^{-1} (\mathbf{M}^T \mathbf{P}^{-*} + \mathbf{P}^{-1} \mathbf{M}) \Lambda^{-1} \mathbf{s}.\end{aligned} \quad (84)$$

We can now use this formula to express the following matrix:

$$\begin{aligned}\Xi &= \mathbb{E}[\phi(\Lambda \mathbf{s}) \mathbf{s}^T \Lambda] = \Lambda^{-1} (\mathbf{P}^{-1} \mathbb{E}[\mathbf{s}^* \mathbf{s}^T] \\ &\quad - \frac{1}{2} (\mathbf{M}^T \mathbf{P}^{-*} + \mathbf{P}^{-1} \mathbf{M}) \mathbb{E}[\mathbf{s} \mathbf{s}^T]) \Lambda = \\ &\quad \Lambda^{-1} \left(\mathbf{P}^{-1} \Sigma^* - \frac{1}{2} (\mathbf{M}^T \mathbf{P}^{-*} + \mathbf{P}^{-1} \mathbf{M}) \Gamma \right) \Lambda = \mathbf{I}_K,\end{aligned} \quad (85)$$

where we have used (83). Now, (45) follows since $\mu_k = \mathbf{e}_k^H \Xi \mathbf{e}_k$.

It is easily seen that, when Σ , Γ , and Λ are replaced, respectively, by their estimates $\tilde{\Sigma}$, $\tilde{\Gamma}$, and $\tilde{\Lambda}$ (assuming that $\tilde{\Sigma}^{-1}$ and $\tilde{\mathbf{P}}^{-1}$ exist), we can follow the same steps to prove (47). Provided that $\hat{\mathbb{E}}[\mathbf{s}^* \mathbf{s}^T]$ is in (85) replaced by $\tilde{\Sigma}^*$ and $\hat{\mathbb{E}}[\mathbf{s} \mathbf{s}^T]$ is replaced by $\tilde{\Gamma}$, the assertion that $\hat{\nu}_k = 1$ follows.

Finally, (46) and (48) readily follow by considering the Wirtinger derivative of (84) by \mathbf{s}^* .

REFERENCES

- [1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. (Independent Component Analysis and Applications Series), Amsterdam, The Netherlands: Elsevier Science, 2010.
- [2] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.
- [3] T. Adali, M. Anderson, and G. S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 18–33, May 2014.
- [4] D. T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. Signal Process.*, vol. 45, no. 7, pp. 1712–1725, Jul. 1997.
- [5] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.
- [6] J.-P. Nadal and N. Parga, "Nonlinear neurons in the low-noise limit: A factorial code maximizes information transfer," *Network: Comput. Neural Syst.*, vol. 5, no. 4, pp. 565–581, 1994.
- [7] J.-F. Cardoso, "Infomax and maximum likelihood for blind source separation," *IEEE Signal Process. Lett.*, vol. 4, no. 4, pp. 112–114, Apr. 1997.
- [8] Z. Koldovský, P. Tichavský, and E. Oja, "Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound," *IEEE Trans. Neural Netw.*, vol. 17, no. 5, pp. 1265–1277, Sep. 2006.
- [9] R. Boscolo, H. Pan, and V. P. Roychowdhury, "Independent component analysis based on nonparametric density estimation," *IEEE Trans. Neural Netw.*, vol. 15, no. 1, pp. 55–65, Jan. 2004.
- [10] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Proc. Neural Inf. Process. Syst.*, 1996, pp. 757–763.
- [11] O. Nobutaka and S. Miyabe, "Auxiliary-function-based independent component analysis for super-gaussian sources," in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*, Berlin Heidelberg, 2010, pp. 165–172.
- [12] A. Hyvärinen, "Fast and robust fixed-point algorithm for independent component analysis," *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, May 1999.
- [13] Z. Koldovský and P. Tichavský, "Gradient algorithms for complex non-Gaussian independent component/vector extraction, question of convergence," *IEEE Trans. Signal Process.*, vol. 67, no. 4, pp. 1050–1064, Feb. 2019.
- [14] 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 Trans. Signal Process.*, vol. 69, pp. 2158–2173, 2021.
- [15] Z. Koldovský, P. Tichavský, and N. Ono, "Orthogonally-constrained extraction of independent non-Gaussian component from non-Gaussian background without ICA," in *Proc. Int. Conf. Latent Variable Anal. Signal Separation*, 2018, pp. 161–170.

- [16] V. Kautský, Z. Koldovský, P. Tichavský, and V. Zaroso, "Cramér–Rao bounds for complex-valued independent component extraction: Determined and piecewise determined mixing models," *IEEE Trans. Signal Process.*, vol. 68, pp. 5230–5243, 2020.
- [17] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
- [18] D.-T. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.
- [19] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.
- [20] P. Tichavský and A. Yeredor, "Fast approximate joint diagonalization incorporating weight matrices," *IEEE Trans. Signal Process.*, vol. 57, no. 3, pp. 878–891, Mar. 2009.
- [21] D. Nion, "A tensor framework for nonunitary joint block diagonalization," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4585–4594, Oct. 2011.
- [22] P. Tichavský and Z. Koldovský, "Algorithms for nonorthogonal approximate joint block-diagonalization," in *Proc. 20th Eur. Signal Process. Conf.*, 2012, pp. 2094–2098.
- [23] D. Lahat and C. Jutten, "Joint independent subspace analysis using second-order statistics," *IEEE Trans. Signal Process.*, vol. 64, no. 18, pp. 4891–4904, Sep. 2016.
- [24] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 1, pp. 70–79, Jan. 2007.
- [25] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop Appl. Signal Process. Audio Acoust.*, 2011, pp. 189–192.
- [26] M. Anderson, G. Fu, R. Phlypo, and T. Adali, "Independent vector analysis: Identification conditions and performance bounds," *IEEE Trans. Signal Process.*, vol. 62, no. 17, pp. 4399–4410, Sep. 2014.
- [27] R. Scheibler and N. Ono, "Independent vector analysis with more microphones than sources," in *Proc. IEEE Workshop Appl. Signal Process. to Audio Acoust.*, 2019, pp. 185–189.
- [28] Y. Li, T. Adali, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Trans. Signal Process.*, vol. 57, no. 10, pp. 3918–3929, Oct. 2009.
- [29] A. Weiss, S. A. Cheema, M. Haardt, and A. Yeredor, "Performance analysis of the Gaussian quasi-maximum likelihood approach for independent vector analysis," *IEEE Trans. Signal Process.*, vol. 66, no. 19, pp. 5000–5013, Oct. 2018.
- [30] W. Kellermann, H. Buchner, and R. Aichner, "Separating convolutive mixtures with TRINICON," in *Proc. IEEE Int. Conf. Audio, Speech Signal Process.*, 2006, pp. 961–964.
- [31] Z. Koldovský, J. Málek, P. Tichavský, Y. Deville, and S. Hosseini, "Blind separation of piecewise stationary non-Gaussian sources," *Signal Process.*, vol. 89, no. 12, pp. 2570–2584, 2009.
- [32] X.-L. Li and T. Adali, "Independent component analysis by entropy bound minimization," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5151–5164, Oct. 2010.
- [33] N. Ono, "Auxiliary-function-based independent vector analysis with power of vector-norm type weighting functions," in *Proc. Asia Pacific Signal Inf. Process. Assoc. Annu. Summit Conf.*, 2012, pp. 1–4.
- [34] 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 Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 9, pp. 1626–1641, Sep. 2016.
- [35] D. Kitamura et al., "Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation," *EURASIP J. Adv. Signal Process.*, vol. 2018, no. 1, pp. 1–25, 2018.
- [36] S. Mogami et al., "Independent low-rank matrix analysis based on time-variant sub-Gaussian source model for determined blind source separation," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 503–518, 2020.
- [37] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.
- [38] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2001.
- [39] P. Smaragdīs, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.
- [40] S. Makino, T.-W. Lee, and H. Sawada, Eds., *Blind Speech Separation*. Berlin, Germany: Springer, 2007.
- [41] E. Vincent, T. Virtanen, and S. Gannot, *Audio Source Separation and Speech Enhancement*, 1st ed. Hoboken, NJ, USA: Wiley Publishing, 2018.
- [42] B. Ehsandoust, M. Babaie-Zadeh, B. Rivet, and C. Jutten, "Blind source separation in nonlinear mixtures: Separability and a basic algorithm," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4339–4352, Aug. 2017.
- [43] Y. Deville, L. Duarte, and S. Hosseini, *Nonlinear Blind Source Separation and Blind Mixture Identification: Methods for Bilinear, Linear-Quadratic and Polynomial Mixtures*. SpringerBriefs in Electrical and Computer Engineering, Berlin, Germany: Springer, 2021.
- [44] T. Taniguchi, N. Ono, A. Kawamura, and S. Sagayama, "An auxiliary-function approach to online independent vector analysis for real-time blind source separation," in *Proc. 4th Joint Workshop Hands-free Speech Commun. Microphone Arrays*, 2014, pp. 107–111.
- [45] S. H. Hsu, T. R. Mullen, T. P. Jung, and G. Cauwenberghs, "Real-time adaptive EEG source separation using online recursive independent component analysis," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 24, no. 3, pp. 309–319, Mar. 2016.
- [46] A. Yeredor, "TV-SOBI: An expansion of SOBI for linearly time-varying mixtures," in *Proc. 4th Int. Symp. Independent Compon. Anal. Blind Source Separation*, 2003, pp. 903–908.
- [47] T. Weisman and A. Yeredor, "Separation of periodically time-varying mixtures using second-order statistics," in *Independent Component Analysis and Blind Signal Separation*, J. Rosca, D. Erdogmus, J. C. Principe, and S. Haykin, Eds. Berlin, Heidelberg: Springer, 2006, pp. 278–285.
- [48] J. Janský, Z. Koldovský, J. Málek, T. Kounovský, and J. Čmejla, "Auxiliary function-based algorithm for blind extraction of a moving speaker," *EURASIP J. Audio, Speech, Music Process.*, vol. 2022, no. 1, pp. 1–16, 2022.
- [49] T. Kim, I. Lee, and T. Lee, "Independent vector analysis: Definition and algorithms," in *Proc. 40th Asilomar Conf. Signals, Syst. Comput.*, 2006, pp. 1393–1396.
- [50] K. Matsuoka and S. Nakashima, "Minimal distortion principle for blind source separation," in *Proc. Int. Conf. Independent Compon. Anal. Signal Separation*, 2001, pp. 722–727.
- [51] Z. Koldovský and F. Nesta, "Performance analysis of source image estimators in blind source separation," *IEEE Trans. Signal Process.*, vol. 65, no. 16, pp. 4166–4176, Aug. 2017.
- [52] K. Žmolíková et al., "Speakerbeam: Speaker aware neural network for target speaker extraction in speech mixtures," *IEEE J. Sel. Topics Signal Process.*, vol. 13, no. 4, pp. 800–814, Aug. 2019.
- [53] Z. Pan, R. Tao, C. Xu, and H. Li, "Muse: Multi-modal target speaker extraction with visual cues," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2021, pp. 6678–6682.
- [54] J. Málek, J. Janský, Z. Koldovský, T. Kounovský, J. Čmejla, and J. Žďánský, "Target speech extraction: Independent vector extraction guided by supervised speaker identification," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 30, pp. 2295–2309, 2022.
- [55] H. Li and T. Adali, "Complex-valued adaptive signal processing using nonlinear functions," *EURASIP J. Adv. Signal Process.*, vol. 2008, pp. 1–9, 2008.
- [56] Z. Koldovský, V. Kautský, T. Kounovský, and J. Čmejla, "Algorithm for independent vector extraction based on semi-time-variant mixing model," in *Proc. 29th Eur. Signal Process. Conf.*, 2021, pp. 865–869.
- [57] I. Lee, T. Kim, and T.-W. Lee, "Fast fixed-point independent vector analysis algorithms for convolutive blind source separation," *Signal Process.*, vol. 87, no. 8, pp. 1859–1871, 2007.
- [58] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Trans. Signal Process.*, vol. 44, no. 10, pp. 2637–2640, Oct. 1996.
- [59] R. A. Usmani, "Inversion of a tridiagonal jacobi matrix," *Linear Algebra Appl.*, vol. 212, pp. 413–414, 1994.
- [60] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution-characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1427–1433, Mar. 2010.
- [61] Z. Koldovský, J. Málek, and J. Janský, "Extraction of independent vector component from underdetermined mixtures through block-wise determined modeling," in *Proc. IEEE Int. Conf. Audio, Speech Signal Process.*, 2019, pp. 7903–7907.



Zbyněk Koldovský (Senior Member, IEEE) was born in Jablonec nad Nisou, Czech Republic, in 1979. He received the M.S. and Ph.D. degrees in mathematical modeling from the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, Czech Republic, in 2002 and 2006, respectively. Since 2020, he has been a full Professor with the Institute of Information Technology and Electronics, Technical University of Liberec, Liberec, Czech Republic, and the Leader of Acoustic Signal Analysis and Processing Group. He is the

Associated Dean for Science, Research and Doctoral Studies with the Faculty of Mechatronics, Informatics and Interdisciplinary Studies. His main research interests include audio signal processing, blind source separation, independent component analysis, and sparse representations. He was the General Co-Chair of the 12th Conference on Latent Variable Analysis and Signal Separation, Liberec, Czech Republic, and as the Technical Co-Chair of the 16th International Workshop on Acoustic Signal Enhancement, Tokyo, Japan. Since 2019, he has been a Member of the IEEE SPS committee Audio and Acoustic Signal Processing.



Petr Tichavský (Senior Member, IEEE) received the Ph.D. degree in theoretical cybernetics and the Research Professor degree from the Czechoslovak Academy of Sciences, Prague, Czech Republic, in 1992 and 2017, respectively. He is currently with the Institute of Information Theory and Automation, Czech Academy of Sciences, Prague, Czechia. He is the author or coauthor of research articles in the area of sinusoidal frequency/frequency-rate estimation, adaptive filtering and tracking of time-varying signal parameters, algorithm-independent bounds on

achievable performance, sensor array processing, independent component analysis, blind source separation, and tensor decompositions. Dr. Tichavský has been a Member of the IEEE SPS Committee Signal Processing Theory and Methods from 2008 to 2011 and since 2016. He was also the General Co-Chair of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP 2011 in Prague. He was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS from 2002 to 2004 and an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 2005 to 2009 and from 2011 to 2016.



Václav Kautský received the master's degree (and started the Ph.D. degree) in applied mathematics from the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University, Prague, Czechia, in 2016. He is currently with the Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University Liberec, Liberec, Czechia. He is author or coauthor of research papers in the area of blind source extraction and algorithm-independent bounds on achievable performance.