

# Dynamic Independent Component/Vector Analysis: Time-Variant Linear Mixtures Separable by Time-Invariant Beamformers

Zbyněk Koldovský , Senior Member, IEEE, Václav Kautský , Student Member, IEEE, Petr Tichavský , Senior Member, IEEE, Jaroslav Čmejla, and Jiří Málek , Member, IEEE

**Abstract**—A novel extension of Independent Component and Independent Vector Analysis for blind extraction/separation of one or several sources from time-varying mixtures is proposed. The mixtures are assumed to be separable source-by-source in series or in parallel based on a recently proposed mixing model that allows for the movements of the desired source while the separating beamformer is time-invariant. The popular FastICA algorithm is extended for these mixtures in one-unit, symmetric and block-deflation variants. The algorithms are derived within a unified framework so that they are applicable in the real-valued as well as complex-valued domains, and jointly to several mixtures, similar to Independent Vector Analysis. Performance analysis of the one-unit algorithm is provided; it shows its asymptotic efficiency under the given mixing and statistical models. Numerical simulations corroborate the validity of the analysis, confirm the usefulness of the algorithms in separation of moving sources, and show the superior speed of convergence and ability to separate super-Gaussian as well as sub-Gaussian signals.

**Index Terms**—Blind Source Separation, Blind Source Extraction, Independent Component Analysis, Independent Vector Analysis, Dynamic Models, Moving Sources.

## I. INTRODUCTION

**I**NDEPENDENTComponent Analysis (ICA) is a popular method proposed for Blind Source Separation (BSS) [1]–[3]. Signals observed on  $d$  sensors are assumed to be linear mixtures

Manuscript received July 22, 2020; revised December 14, 2020 and March 1, 2021; accepted March 20, 2021. Date of publication March 26, 2021; date of current version April 28, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Pia Addabbo. This work was supported in part by The Czech Science Foundation through Project 20-17720S, and by the department of the Navy, Office of Naval Research Global, through Project N62909-19-1-2105. Matlab implementations of FastDIVA and of the example presented in Section I are available at <https://asap.ite.tul.cz/downloads/ice/>. (Corresponding author: Zbynek Koldovsky.)

Zbyněk Koldovský, Jaroslav Čmejla, and Jiří Málek are with the Acoustic Signal Analysis and Processing Group, Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, 246117 Liberec, Czechia (e-mail: zbynek.koldovsky@tul.cz; jaroslav.cmejla@tul.cz; jiri.malek@tul.cz).

Václav Kautský is with the Acoustic Signal Analysis and Processing Group, Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec, 246117 Liberec, Czechia, and also with the Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, 166 36 Prague, Czechia (e-mail: kautsvac@jfifi.cvut.cz).

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

Digital Object Identifier 10.1109/TSP.2021.3068626

of  $d$  “original” signals, which are mutually independent in the statistical sense. The linear mixing model is given by

$$\mathbf{x}^n = \mathbf{A}^n \mathbf{s}^n, \quad (1)$$

where  $n = 1, \dots, N$  is the sample index,  $\mathbf{x}^n$  is a  $d \times 1$  vector of the observed mixed signals at time  $n$ ;  $\mathbf{A}^n$  is a  $d \times d$  non-singular mixing matrix; and  $\mathbf{s}^n$  is a  $d \times 1$  vector of the original independent signals. Since  $\mathbf{A}^n$  are square and non-singular, the model is referred to as *determined*. We speak about the *static* mixing case when  $\mathbf{A}^n$  is constant over  $n$ . ICA can be formulated as to estimate  $(\mathbf{A}^n)^{-1}$  through finding square de-mixing matrices  $\mathbf{W}^n$  such that the signals  $\mathbf{W}^n \mathbf{x}^n$  are as independent as possible. It is the indeterminacy of BSS (as well as of ICA) that the order and scales of  $\mathbf{s}^n$  cannot be retrieved without additional information.

The determined static formulation has become very popular mainly due to its mathematical tractability and wide applicability. The problem has been deeply studied and, currently, ICA and its extension to joint separation of several mixtures (data sets) such as Independent Vector Analysis (IVA), have matured to a large extent [4]–[8]. For most recent contributions to the area see, e.g., [9], [10].

In many applications, however, it is necessary to consider the time-variant mixing model, which we will refer to as *dynamic*. For example, in audio or biomedical applications it happens that the mixing environment is changing in time, sources are moving, some new sources can appear randomly in time and some other may disappear. There is therefore a need to estimate the mixing/de-mixing matrix in an adaptive manner, respecting the dynamic nature of the data.

The determined mixing model (1) with  $\mathbf{A}^n$  dependent on  $n$  can capture a very wide class of dynamic mixtures. However, the lack of information (the number of samples  $N$  is proportional to the number of unknown parameters  $Nd^2$ ) and the random order of the separated signals at any time instant pose crucial problems. Current extensions of ICA and IVA towards the dynamic model are therefore based on more or less strictly formulated assumptions that the changes of the mixing parameters are somewhat slow and continuous. A standard way is that estimation methods for the static case are converted into adaptive algorithms along the lines of the least-mean-squares (LMS) or recursive-least-squares (RLS) algorithms [11]. To this

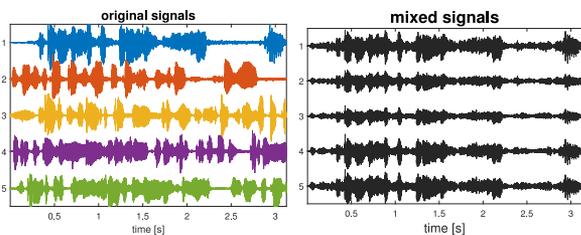


Fig. 1. LEFT: Five independent speech signals, each 50 000 samples long, sampled at 16 kHz. RIGHT: Instantaneous mixture of the signals where the first signal is linearly moving while signals 2 through 5 are static. Signal 1 is dominating the mixture (multiplied by factor 5).

end, various sequential [12], recursive [13], [14], Bayesian [15] or other online approaches have been proposed. Particularly popular adaptive methods are based on the Natural Gradient algorithm [16], [17]; see, e.g., [14], [18]–[21]. In biomedical application, an Online Recursive Independent Component Analysis (ORICA) was proposed in [14], [22]. The latter paper is remarkable because it presents a real-world application of the algorithm in high-density (64 channel) EEG.

The approach that we present here is conceptually different. Basically, it is off-line, despite it allows to handle time-varying mixtures to certain extent. It comes from the recently proposed blind source extraction (BSE) model denoted as CSV (Constant Separating Vector) where the mixing parameters related to the source of interest (SOI) can be varying in time while the de-mixing parameters are time-invariant [23], [24]. CSV allows for the SOI movements throughout the exposed data. The Cramér-Rao analysis has been done in [25]. It points to appealing properties of CSV in terms of the achievable extraction accuracy compared to the sequentially applied ICA. On-line version of the proposed approach is possible as well, because we can think about allowing the “constant” separating vector to be progressively updated in time.

This paper brings two major contributions. First, we extend CSV to separation of more than one source, by which we introduce so-called *CSV-separable* mixtures. Briefly, the CSV-separable mixtures are defined as such that can be separated source-by-source in series or in parallel based on the CSV. In fact, the formulation of ICA/IVA on CSV-separable mixtures is a natural extension of the static ICA/IVA to the special class of dynamic mixtures. It provides a novel tool for off-line exploratory data analysis and is also useful in online data processing, as we demonstrate in the experimental section. Second, we propose the FastDIVA algorithm (Fast Dynamic Independent Vector Analysis) as a new method for ICA/IVA on CSV-separable mixtures. In fact, FastDIVA is a successor of the famous FastICA [26] and FastIVA [27] as it involves these methods as special cases and is proposed in three variants: one-unit, symmetric and block-deflation. To motivate, we provide the following example.

Consider five speech signals shown in Fig. 1 (left). Their instantaneous<sup>1</sup> mixture, shown in Fig. 1 (right), is generated

<sup>1</sup>Note that this mixture is not convolutive as is typical to real-world acoustic signal mixing; we consider the simpler instantaneous case for demonstration purposes.

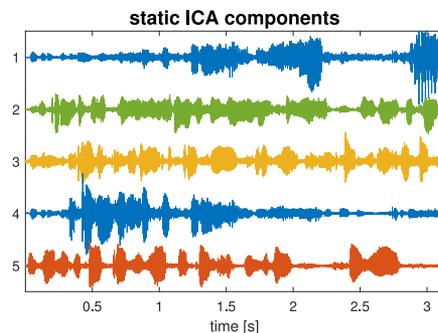


Fig. 2. Independent components extracted from the signal mixture shown in the right part of Fig. 1 by symmetric FastICA.

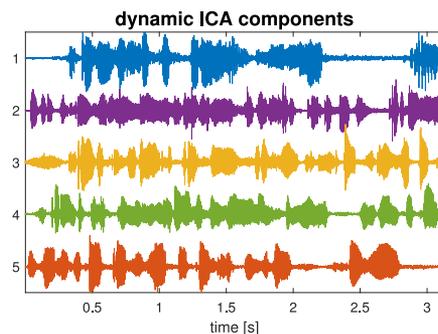


Fig. 3. Independent components separated from the dynamic mixture in Fig. 1 (right) by block-deflation FastDIVA set to 5 blocks, each of length  $10^4$  samples.

so that signals 2 through 5 are static, mixed into 5 channels with fixed random mixing coefficients, while signal 1 (i.e., its virtual source) is moving: The first column of  $\mathbf{A}^n$ , denoted as  $\mathbf{a}^n$ , is linearly progressing from  $\mathbf{a}$  to  $\mathbf{b}$  according to  $\mathbf{a}^n = (N - n + 1)/N\mathbf{a} + (n - 1)/N\mathbf{b}$ ;  $\mathbf{a}$  and  $\mathbf{b}$  are random column vectors that make an angle of  $20^\circ$ . The other columns of  $\mathbf{A}^n$  are constant over  $n$ . Signal 1 is amplified by factor 5 in order to accentuate it in the mixture.

Fig. 2 shows typical components obtained by a conventional ICA algorithm (symmetric FastICA [26]) when applied to this mixture. The order of components is random, which is due to the inherent ambiguity of BSS. By visual inspection, components 2 and 5 correspond to the original signals 5 and 2, respectively, up to scales and signs. Component 3 corresponds to the original signal 3 up to a certain more significant residual interference. Components 4 and 1 consist of the beginning and end parts of the original signal 1, respectively. This is caused by the movement of the corresponding (virtual) source. The original signal 4 is not extracted as a separated component; it appears as a residual within component 4. Note that this situation cannot be improved by extracting one more component because the static ICA (de-)mixing model assumes square (de-)mixing matrix.

Fig. 3 shows components that have been separated by block-deflation FastDIVA assuming CSV-separable mixing model with 5 blocks. The components correspond with the original signals up to a random order, which is 1, 4, 3, 5, 2, and a reasonable statistical error. Not only does the algorithm extract the moving

signal as one component, that is, without the need for collecting it from several components whose order is unknown. It also separates original signal  $\mathbf{x}$  with a high degree of precision, as compared to symmetric FastICA.

This paper is organized as follows. The CSV mixing model is revised and extended to the CSV-separable mixtures in Section II. The FastDIVA algorithm is proposed in Section III and its one-unit version is analyzed in Section IV. Numerical experiments and comparisons with FastDIVA in off-line and on-line tests are provided in Section V; and Section VI concludes the article.

### I. 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.,  $[\mathbf{1}; \mathbf{g}] = [1, \mathbf{g}^T]^T$ . The statistical models of signals considered in this paper assume that each sample is independently drawn from a distribution; inter-sample dependencies are not modeled. Therefore, we use symbolic notation where samples having the same distribution are represented by random (vector) variables.  $E[\cdot]$  stands for the expectation value of the argument, and  $\hat{E}[\cdot]$  is the average value of the argument taken over all of its available samples. The letters  $k$ ,  $t$ , and  $i$  are used as integer indices of dataset, block, and source, respectively; index omission will always be announced in the text.  $\{\cdot\}_k$  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  $\langle a_t \rangle_t$ .

We will consider complex-valued signals and parameters; however, the conclusions of this work are valid for the real-valued case as well.

## II. PROBLEM FORMULATION

For practical reasons, we turn from (1) to mixtures that are block-wise static and, also, extend our considerations to multiple datasets as in IVA and in other joint BSS problems [28]–[32].

Let  $N$  samples of signals be observed through  $d$  sensors in  $K$  datasets, and let the samples be divided into  $T \geq 1$  time-intervals called blocks. For the sake of simplicity, let the blocks have the same length  $N_b$ , and  $N = T \cdot N_b$ . From now on, we will consider the block-wise varying mixing model

$$\mathbf{x}^{k,t} = \mathbf{A}^{k,t} \mathbf{s}^{k,t}, \quad (2)$$

where  $k = 1, \dots, K$  is the dataset index;  $t = 1, \dots, T$  is the block index;  $\mathbf{A}^{k,t}$  is a  $d \times d$  non-singular mixing matrix; and by  $\mathbf{s}^{k,t} = [s_1^{k,t}, \dots, s_d^{k,t}]^T$  we denote *independent* random variables representing unknown original signals. Without any loss of generality, let all the signals have zero mean values. Samples of signals within the blocks are assumed identically and independently distributed (i.i.d.). For example, each sample of the  $i$ th original signal within the  $t$ th block and the  $k$ th dataset is assumed to be an independent realization of the random variable  $s_i^{k,t}$ . By taking into account the ambiguities, the BSS task can be, in general, formulated as follows.

Find de-mixing matrices  $\mathbf{W}^{k,t}$  such that  $\mathbf{W}^{k,t} \mathbf{x}^{k,t}$  are equal to  $\mathbf{s}^{k,t}$  up to their original scales and phase. The order of the separated signals can be different from the original one; however, it is desirable for it to be the same in all datasets and blocks.

For  $T = 1$ , we have the *static* case considered by the conventional ICA and IVA. In ICA, the datasets are separated independently; this approach, however, brings random permutation of separated signals in the datasets, the permutation problem [33]. In IVA, components are separated as “vectors” where the  $i$ th vector component is defined as  $\mathbf{s}_i^t = [s_i^{1,t}, \dots, s_i^{K,t}]^T$ ,  $i = 1, \dots, d$  [34].

We are mainly interested in the dynamic case of  $T > 1$ , where the mixing parameters (matrices) can be varying from block to block. ICA and IVA can be used when  $T > 1$  by applying them separately on blocks. However, this approach does not guarantee the same order, i.e., continuity of the separated signals over the blocks, a phenomenon caused by the uncertainty of signal order similar to the permutation problem; we refer to it as *the discontinuity problem*. Also, there are too many parameters to be estimated, which potentially leads to a loss in separation accuracy.

What we basically do, in this paper, is applying a deflation or symmetric manner of blind source separation as in deflation or symmetric FastICA when  $T = 1$  [26]. It means that we wish to separate the signal components one by one or in parallel. Therefore, the primary problem to be solved is the blind extraction of one component. For  $T = 1$ , this is solved through so-called Independent Component or Independent Vector Extraction (ICE/IVE).

In ICE/IVE, it is reflected that if only the SOI should be extracted, only the corresponding column of  $\mathbf{A}^{k,t}$  and the corresponding row of  $(\mathbf{A}^{k,t})^{-1}$  need to be taken into account in the mixing model parameterization. The other columns of  $\mathbf{A}^{k,t}$  need not be estimated, only their corresponding subspace should be identified. The parameterization chosen in [35] ensures this.

Owing to the indeterminacy of order in BSS<sup>2</sup>, we can assume that the SOI corresponds to  $s_1^{k,t}$ . Hence, according to [35],  $\mathbf{A}^{k,t}$  in (2) can be parameterized by

$$\mathbf{A}_{\text{BSE}}^{k,t} = \begin{pmatrix} \mathbf{a}^{k,t} & \mathbf{Q}^{k,t} \end{pmatrix} = \begin{pmatrix} \gamma^{k,t} & (\mathbf{h}^{k,t})^H \\ \mathbf{g}^{k,t} & \frac{1}{\gamma^{k,t}} (\mathbf{g}^{k,t} (\mathbf{h}^{k,t})^H - \mathbf{I}_{d-1}) \end{pmatrix} \quad (3)$$

where  $\mathbf{a}^{k,t} = [\gamma^{k,t}; \mathbf{g}^{k,t}]$  is called the *mixing vector* corresponding to the first column of  $\mathbf{A}^{k,t}$ ;  $\mathbf{I}_d$  denotes the  $d \times d$  identity matrix;  $(\mathbf{w}^{k,t})^H$  denotes the first row of  $\mathbf{W}_{\text{BSE}}^{k,t} = (\mathbf{A}_{\text{BSE}}^{k,t})^{-1}$ ;  $\mathbf{w}^{k,t}$  stands for the beamformer on which output is the extracted signal, i.e.,  $s_1^{k,t} = (\mathbf{w}^{k,t})^H \mathbf{x}^{k,t}$ ; we will call it the *separating vector*. It holds that

$$\mathbf{W}_{\text{BSE}}^{k,t} = \begin{pmatrix} (\mathbf{w}^{k,t})^H \\ \mathbf{B}^{k,t} \end{pmatrix} = \begin{pmatrix} (\beta^{k,t})^* & (\mathbf{h}^{k,t})^H \\ \mathbf{g}^{k,t} & -\gamma^{k,t} \mathbf{I}_{d-1} \end{pmatrix}, \quad (4)$$

where  $\mathbf{w}^{k,t} = [\beta^{k,t}; \mathbf{h}^{k,t}]$ ,  $\mathbf{B}^{k,t} = [\mathbf{g}^{k,t}, -\gamma^{k,t} \mathbf{I}_{d-1}]$  satisfies the condition  $\mathbf{B}^{k,t} \mathbf{a}^{k,t} = \mathbf{0}$  (the blocking matrix [36]). Since

<sup>2</sup>In fact, any knowledge about the SOI (e.g., an initial guess) must be available to determine it; see Section II.B in [35]

$\mathbf{W}_{\text{BSE}}^{k,t} \mathbf{A}_{\text{BSE}}^{k,t} = \mathbf{I}_d$ ,  $\mathbf{a}^{k,t}$  and  $\mathbf{w}^{k,t}$  are linked through the so-called *distortionless constraint*  $(\mathbf{w}^{k,t})^H \mathbf{a}^{k,t} = 1$ , which can also be written as

$$(\beta^{k,t})^* \gamma^{k,t} = 1 - (\mathbf{h}^{k,t})^H \mathbf{g}^{k,t}. \quad (5)$$

The subspace of the other signals, referred to as *background*, is generated by  $\mathbf{z}^{k,t} = \mathbf{B}^{k,t} \mathbf{x}^{k,t}$ . The fact that  $\mathbf{B}^{k,t} \mathbf{a}^{k,t} = \mathbf{0}$  guarantees that  $\mathbf{z}^{k,t}$  span the same subspace as  $s_2^{k,t}, \dots, s_d^{k,t}$ .

The approach where ICE/IVE is applied separately to each block when  $T > 1$  has the same drawbacks as mentioned above. In [23], [25], we have distinguished two simplifications, or two particular separation models.

- Constant mixing vector (CMV)

$$\mathbf{a}^{k,1} = \mathbf{a}^{k,2} = \dots = \mathbf{a}^{k,T} = \mathbf{a}^k \quad (6)$$

- Constant separating vector (CSV)

$$\mathbf{w}^{k,1} = \mathbf{w}^{k,2} = \dots = \mathbf{w}^{k,T} = \mathbf{w}^k \quad (7)$$

Each of the two models ensures that the components in different blocks are not permuted randomly, thus, the discontinuity problem is avoided. The former model might be more suitable when the sources are not moving, but the background is non-stationary, or that there is low signal-to-interference and noise ratio. The latter models appears to be useful when the SOI is moving, which is of a greater interest. Therefore, we deal with the CSV model, in this paper, and generalize it to separation of  $r \geq 1$  sources.

We introduce the notion of *CSV-separable mixtures* through the following conditions:

- C1) All  $r$  sources to be separated obey CSV, which means that the first  $r$  rows of  $(\mathbf{A}^{k,1})^{-1}, \dots, (\mathbf{A}^{k,T})^{-1}$  in (2) are constant over  $t$ .
- C2) For each  $i = 1, \dots, r$ , the  $i$ th source obeys CSV in a reduced mixture where sources  $1, \dots, i-1$  have been subtracted.

Some properties readily follow. For  $r = d$ , all rows of the inverse matrices are assumed constant in (C1), which means that the mixtures obeying (C1) are static when  $r = d$ . For  $r = 1$ , (C1) and (C2) coincide with the CSV model for one source.

Validity of the conditions (C1) and (C2) has to be assumed. Their usefulness was already shown in Section I and will be supported by additional examples in Section V. In Section III-G, we propose the symmetric and block-deflation separation schemes,<sup>3</sup> which can be used to separate  $r$  sources from mixtures obeying (C1) and (C2), respectively.

### III. PROPOSED ALGORITHM

The detailed derivations of one-unit, symmetric and block-deflation FastDIVA are provided in this Section. We begin with the one-unit variant, which solves the BSE problem based on the CSV mixing model.

<sup>3</sup>The symmetric, deflation and block-deflation separation schemes are here applied together with the one-unit FastDIVA algorithm, however, they can be applied with other BSE algorithms.

#### A. Statistical Model

To simplify notation, for now, we will omit the subscript “1” in  $\mathbf{s}_1^t$ , i.e.,  $\mathbf{s}^t = [s^{1,t}, \dots, s^{K,t}]^T$ . Let the probability density function (pdf) of  $\mathbf{s}^t$  be  $p(\mathbf{s}^t)$ . Note that this pdf is, in general, dependent on  $t$ ; we do not write this explicitly, for simplicity. Next, let  $p_{\mathbf{z}^{k,t}}(\mathbf{z}^{k,t})$  denote the pdf of  $\mathbf{z}^{k,t}$ . Although there can also be dependencies between background signals from different datasets, we neglect them to simplify the statistical model of the background<sup>4</sup>.

Considering the structure of the de-mixing matrix (4) with the CSV assumption (7), using the independence between the SOI and the background, and taking into account the fact that samples are independently distributed, we get the joint pdf for one sample of the observed signals in the  $t$ th block in the form

$$p_{\mathbf{x}^{k,t}}(\{\mathbf{x}^{k,t}\}_k) = p(\{(\mathbf{w}^k)^H \mathbf{x}^{k,t}\}_k) \times \prod_{k=1}^K p_{\mathbf{z}^{k,t}}(\mathbf{B}^{k,t} \mathbf{x}^{k,t}) |\det \mathbf{W}_{\text{BSE}}^{k,t}|^2. \quad (8)$$

Note that the square of the absolute value of determinant is necessary due to the transformation of densities of the complex-valued random variables (the exponent equals one in the real-valued case). The determinant can be expressed by using Eq. (15) in [35], which gives, together with (7),  $|\det \mathbf{W}_{\text{BSE}}^{k,t}|^2 = |\gamma^{k,t}|^{2(d-2)}$ . The pdf of all  $N$  samples is equal to  $\prod_{t=1}^T p_{\mathbf{x}^{k,t}}(\{\mathbf{x}^{k,t}\}_k)^{N_b}$ , so the log-likelihood function divided by  $N$  can be expressed as

$$\begin{aligned} \mathcal{L}(\{\mathbf{w}^k, \mathbf{a}^{k,t}\}_{k,t}) &= \left\langle \hat{\mathbb{E}} [\log p(\{(\mathbf{w}^k)^H \mathbf{x}^{k,t}\}_k)] \right. \\ &\quad \left. + \sum_{k=1}^K \hat{\mathbb{E}} [p(\mathbf{B}^{k,t} \mathbf{x}^{k,t})] + (d-2) \sum_{k=1}^K \log |\gamma^{k,t}|^2 \right\rangle_t. \quad (9) \end{aligned}$$

#### B. Contrast Function

Finding the appropriate maximum of (9) provides the maximum likelihood estimate of the parameter vectors. However, (9) must be replaced by a valid contrast function because of the unknown pdfs  $p(\mathbf{s}^t)$  and  $p_{\mathbf{z}^{k,t}}(\mathbf{z}^{k,t})$ , which have to be replaced by suitable model densities. In the static case, the model pdfs of the SOI can be scaled to unit variance since there is the scaling ambiguity [26]. However, in the dynamic case, the variance of signals can be changing from block to block which must be taken into account. Therefore, the appropriate surrogate for  $p(\cdot)$  is<sup>5</sup> [38], [39]

$$p(\mathbf{s}^t) \approx f\left(\left\{\frac{s^{k,t}}{\hat{\sigma}^{k,t}}\right\}_k\right) \left(\prod_{k=1}^K \hat{\sigma}^{k,t}\right)^{-2}, \quad (10)$$

where  $f(\cdot)$  should be a suitable normalized non-Gaussian pdf, and  $(\hat{\sigma}^{k,t})^2$  is the sample-based variance of the estimate of  $s^{k,t}$ .

<sup>4</sup>This simplification typically brings a suboptimal performance of BSE as compared to BSS [25], [37]

<sup>5</sup>Note that the square power in (10) is necessary due to considering the complex-valued problem; it would equal one in the real-valued case.

It holds that

$$\hat{\sigma}^{k,t} = \sqrt{(\mathbf{w}^k)^H \widehat{\mathbf{C}}^{k,t} \mathbf{w}^k}, \quad (11)$$

where  $\widehat{\mathbf{C}}^{k,t} = \widehat{\mathbb{E}}[\mathbf{x}^{k,t}(\mathbf{x}^{k,t})^H]$  is the sample-based covariance matrix of  $\mathbf{x}^{k,t}$ ;  $\hat{\sigma}^{k,t}$  is, in fact, a function of  $\mathbf{w}^k$ .

Note that  $f(\cdot)$  could be dependent on  $t$ . However, since there is usually little information about the true pdf, we simplify our considerations by assuming that  $f(\cdot)$  is independent of  $t$ .

The unknown  $p_{\mathbf{z}^{k,t}}(\mathbf{z}^{k,t})$  can be replaced by the zero mean circular<sup>6</sup> Gaussian pdf  $\mathcal{CN}(0, \mathbf{C}_{\mathbf{z}^{k,t}})$ , where  $\mathbf{C}_{\mathbf{z}^{k,t}} = \mathbb{E}[\mathbf{z}^{k,t}(\mathbf{z}^{k,t})^H]$  is the covariance matrix of the background signals; see, e.g., [35] for the justification of this choice.  $\mathbf{C}_{\mathbf{z}^{k,t}}$  is an unknown nuisance parameter, which will later be replaced by its sample-based estimate. By putting the model densities into (9), a practical contrast function for estimating the model parameters takes on the form

$$\begin{aligned} \mathcal{C}(\{\mathbf{w}^k, \mathbf{a}^{k,t}\}_{k,t}) &= \left\langle \widehat{\mathbb{E}} \left[ \log f \left( \left\{ \frac{\hat{s}^{k,t}}{\hat{\sigma}^{k,t}} \right\}_k \right) \right] \right. \\ &- \sum_{k=1}^K \log(\hat{\sigma}^{k,t})^2 - \sum_{k=1}^K \widehat{\mathbb{E}} [(\hat{\mathbf{z}}^{k,t})^H (\mathbf{C}_{\mathbf{z}^{k,t}})^{-1} \hat{\mathbf{z}}^{k,t}] \\ &\left. + (d-2) \sum_{k=1}^K \log |\gamma^{k,t}|^2 \right\rangle + \text{const.}, \quad (12) \end{aligned}$$

where  $\hat{s}^{k,t} = (\mathbf{w}^k)^H \mathbf{x}^{k,t}$ , and  $\hat{\mathbf{z}}^{k,t} = \mathbf{B}^{k,t} \mathbf{x}^{k,t}$ . The remaining constant term is independent of the mixing model parameters. For  $K=1$  and  $T=1$ , the indices  $k$  and  $t$  can be omitted, and (12) is simplified to<sup>7</sup>

$$\begin{aligned} \mathcal{C}^{1,1}(\mathbf{w}, \mathbf{a}) &= \widehat{\mathbb{E}} \left[ \log f \left( \frac{\hat{s}}{\hat{\sigma}} \right) \right] - \log \hat{\sigma}^2 - \widehat{\mathbb{E}} [\hat{\mathbf{z}}^H \mathbf{C}_{\mathbf{z}}^{-1} \hat{\mathbf{z}}] \\ &+ (d-2) \log |\gamma|^2 + \text{const.} \quad (13) \end{aligned}$$

### C. Orthogonal Constraints

The above contrast functions can have many spurious extremes. It may occur that the parameter vectors  $\mathbf{a}^{k,t}$ ,  $t=1, \dots, T$ , and  $\mathbf{w}^k$  do not correspond to the same signal. Therefore, a reliable link between the separating and mixing vectors has to be established. To this end, the orthogonal constraint (OGC) appears to be convenient. Since  $s^{k,t}$  and  $\mathbf{z}^{k,t}$  are independent and, therefore, also uncorrelated, the OGC requires that subspace generated by samples of  $\hat{s}^{k,t}$  is orthogonal to the subspace of  $\hat{\mathbf{z}}^{k,t}$ . Also, (5) must be satisfied. The mixing vectors are then linked with the separating vector through [35]

$$\mathbf{a}^{k,t} = \frac{\widehat{\mathbf{C}}^{k,t} \mathbf{w}^k}{(\mathbf{w}^k)^H \widehat{\mathbf{C}}^{k,t} \mathbf{w}^k}. \quad (14)$$

<sup>6</sup>Noncircular Gaussian pdf could be considered as well, especially, if the background signals are assumed to involve noncircular sources. In Appendix A, it will be shown that the assumption of circularity causes that the Hessian matrix  $\mathbf{H}_1$ , defined later in (30), has rank 1, which significantly simplifies the Newton-Raphson update given by (38).

<sup>7</sup>The reader can compare (13) with 19 in [35]. The contrast functions differ in that (13) involves  $\hat{\sigma}^2$ ; therefore, it contains the normalization inside the argument of  $f(\cdot)$  and an additional second term.

Equivalently,  $\mathbf{w}^k$  can be expressed as the dependent variable as

$$\mathbf{w}^k = \frac{(\widehat{\mathbf{C}}^{k,t})^{-1} \mathbf{a}^{k,t}}{(\mathbf{a}^{k,t})^H (\widehat{\mathbf{C}}^{k,t})^{-1} \mathbf{a}^{k,t}}. \quad (15)$$

### D. Relationship to Optimum Beamformers

The analytic expression (15) corresponds to the minimum power distortionless beamformer (MPDR) steered in the direction determined by the mixing vector  $\mathbf{a}^{k,t}$  when the covariance of data is given by  $\widehat{\mathbf{C}}^{k,t}$ . MPDR is an optimum beamformer known in array processing theory as the solution of [40]

$$\mathbf{w}^k = \arg \min_{\mathbf{w}} \mathbf{w}^H \widehat{\mathbf{C}}^{k,t} \mathbf{w} \quad \text{w.r.t.} \quad \mathbf{w}^H \mathbf{a}^{k,t} = 1. \quad (16)$$

The orthogonally constrained BSE algorithms can, in the static case of  $T=1$ , be viewed as blind MPDR beamformers seeking in the direction of  $\mathbf{a}^{k,t}$ , for a fixed  $t$ , such that the MPDR output is independent of the orthogonal (background) subspace [41].

In the CSV model, (15) and, thus, (16) should be satisfied simultaneously for all  $t=1, \dots, T$ , which imposes  $T$  conditions on one separating vector  $\mathbf{w}^k$ . It is therefore more practical to impose the OGC through (14) rather than through (15) when  $T > 1$ .

In order to interpret the block-independent separating vector in CSV, note that the true mixing and separating vectors satisfy (15) when  $N_b \rightarrow +\infty$ , that is, with  $\widehat{\mathbf{C}}^{k,t}$  replaced by  $\mathbf{C}^{k,t}$ . Hence, the true parameter vectors satisfy

$$\mathbf{w}^k = \arg \min_{\mathbf{w}} \mathbf{w}^H \mathbf{C}^{k,t} \mathbf{w} \quad \text{w.r.t.} \quad \mathbf{w}^H \mathbf{a}^{k,t} = 1 \quad (17)$$

for all  $t=1, \dots, T$ . It follows that they also obey

$$\mathbf{w}^k = \arg \min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}^k \mathbf{w} \quad \text{w.r.t.} \quad \mathbf{w}^H \mathbf{\Lambda}^k = \mathbf{1}, \quad (18)$$

where  $\mathbf{R}^k = \sum_{t=1}^T \mathbf{C}^{k,t}$ ,  $\mathbf{\Lambda}^k = [\mathbf{a}^{k,1} \dots \mathbf{a}^{k,T}]$ , and  $\mathbf{1}$  is the  $T \times 1$  vector of ones. The solution of (18) is known as the linearly constrained minimum power beamformer (LCMP) [40]. We conclude the connection between CSV and LCMP as follows:

For  $N_b \rightarrow +\infty$ , the CSV mixing model ensures that the LCMP beamformer steered in the directions given by the true mixing vectors (determining locations of the SOI during its movement)  $\mathbf{a}^{k,1}, \dots, \mathbf{a}^{k,T}$  exists such that it extracts the SOI from the mixed signals perfectly.

### E. Approximate Newton-Raphson Algorithm

The algorithm proposed here aims at finding a maximum of (12) subject to the parameter vectors  $\mathbf{w}^k$ ,  $k=1, \dots, K$ , under the OGC (14). For the sake of clarity, the contrast function to be maximized is

$$\mathcal{C}_{\text{OG}}(\{\mathbf{w}^k\}_k) = \mathcal{C} \left( \left\{ \mathbf{w}^k, \frac{\widehat{\mathbf{C}}^{k,t} \mathbf{w}^k}{(\mathbf{w}^k)^H \widehat{\mathbf{C}}^{k,t} \mathbf{w}^k} \right\}_{k,t} \right). \quad (19)$$

We follow the complex-valued Newton-Raphson optimization approach using the Wirtinger calculus [42]. This entails the computation of the gradient and the second-order derivatives of  $\mathcal{C}_{\text{OG}}$ . To simplify the exposition, the derivations here will be done as if  $T=1$  and  $K=1$  (the indices  $t$  and  $k$  will be

omitted); the result for  $T \geq 1$  and  $K \geq 1$  will readily follow. Thus, we now compute the derivatives of the four terms in (13) when  $\mathbf{a} = \frac{\hat{\mathbf{C}}_{\mathbf{w}}}{\mathbf{w}^H \hat{\mathbf{C}}_{\mathbf{w}}}$ .

To compute the gradient, we use results from [35] and the following identities

$$\frac{\partial}{\partial \mathbf{w}^H} \hat{s} = \frac{\partial}{\partial \mathbf{w}^H} \mathbf{w}^H \mathbf{x} = \mathbf{x}, \quad (20)$$

$$\frac{\partial}{\partial \mathbf{w}^H} \frac{1}{\hat{\sigma}} = \frac{\partial}{\partial \mathbf{w}^H} \frac{1}{\sqrt{\mathbf{w}^H \hat{\mathbf{C}}_{\mathbf{w}}}} = -\frac{\mathbf{a}}{2\hat{\sigma}}, \quad (21)$$

$$\frac{\partial}{\partial \mathbf{w}^H} \log \hat{\sigma}^2 = \frac{\partial}{\partial \mathbf{w}^H} \log \hat{\mathbf{w}}^H \hat{\mathbf{C}}_{\mathbf{w}} = \mathbf{a}. \quad (22)$$

The derivative of the first term in (13) reads

$$\frac{\partial}{\partial \mathbf{w}^H} \hat{\mathbb{E}} \left[ \log f \left( \frac{\hat{s}}{\hat{\sigma}} \right) \right] = -\hat{\mathbb{E}} \left[ \phi \left( \frac{\hat{s}}{\hat{\sigma}} \right) \frac{\mathbf{x}}{\hat{\sigma}} \right] + \Re\{\hat{\nu}\} \mathbf{a}, \quad (23)$$

where  $\hat{\nu}$  is the sample-based estimate of

$$\nu = \mathbb{E} \left[ \phi \left( \frac{s}{\sigma} \right) \frac{s}{\sigma} \right], \quad (24)$$

$\Re\{\cdot\}$  denotes the real part of the argument, and

$$\phi(s) = -\frac{\partial}{\partial s^*} \log f(s) \quad (25)$$

is the score function corresponding to the model density  $f(\cdot)$ . The derivative of the second term in (13) follows directly from (22). The derivatives of the third and fourth terms are simplified to

$$\frac{\partial}{\partial \mathbf{w}^H} \left\{ -\hat{\mathbb{E}} [\hat{\mathbf{z}}^H \mathbf{C}_{\mathbf{z}}^{-1} \hat{\mathbf{z}}] + (d-2) \log |\gamma|^2 \right\} = \mathbf{a}, \quad (26)$$

when  $\mathbf{C}_{\mathbf{z}}$  is (after taking the derivative) replaced by  $\hat{\mathbf{C}}_{\mathbf{z}}$  as shown in Appendix C in [35]. Hence, terms 2 through 4 in (13) do not contribute to the gradient as their derivatives finally boil down to zero,<sup>8</sup>. The gradient of (19) for  $T = K = 1$  is thus equal to (23), i.e.,

$$\frac{\partial}{\partial \mathbf{w}^H} \mathcal{C}_{\text{OG}}(\mathbf{w}) = \Re\{\hat{\nu}\} \mathbf{a} - \hat{\mathbb{E}} \left[ \phi \left( \frac{\hat{s}}{\hat{\sigma}} \right) \frac{\mathbf{x}}{\hat{\sigma}} \right]. \quad (27)$$

Now, consider  $N \rightarrow +\infty$  and  $\mathbf{w}$  being the true separating vector; if this is the case, (27) is equal to

$$\frac{\partial}{\partial \mathbf{w}^H} \mathcal{C}_{\text{OG}}(\mathbf{w}) = (\Re\{\nu\} - \nu) \mathbf{a}. \quad (28)$$

It follows that the true separating vector is the stationary point of  $\mathcal{C}_{\text{OG}}(\mathbf{w})$  only if  $\Re\{\nu\} = \nu$ . If  $f(\cdot) = p(\cdot)$  then  $\nu = 1$ , and the condition  $\Re\{\nu\} = \nu$  is satisfied. However, this equality does not hold for general  $f(\cdot)$ , so finding the stationary point of  $\mathcal{C}_{\text{OG}}(\mathbf{w})$  need not yield a consistent estimate of the separating vector.

To solve this problem, note that  $f(\cdot)$  does not appear explicitly in (27). We can therefore consider a replacement of  $f(\cdot)$  by its “normalized” variant such that the new score function is  $\hat{\nu}^{-1} \phi(\cdot)$ ,

<sup>8</sup>It follows that BSE methods based on maximizing the non-Gaussianity of the SOI [2], [26], [43] in fact, inherently assume that the background is circular Gaussian with unknown covariance.

and the new  $\nu$  is equal to one. Then, we introduce a modified gradient (27) as

$$\nabla = \mathbf{a} - \hat{\nu}^{-1} \hat{\mathbb{E}} \left[ \phi \left( \frac{\hat{s}}{\hat{\sigma}} \right) \frac{\mathbf{x}}{\hat{\sigma}} \right]. \quad (29)$$

After this modification, the  $\mathbf{w}$  such that  $\nabla = \mathbf{0}$  is a consistent estimate of the true separating vector.

Now, we investigate the second-order derivatives of (19), that is, the derivatives of (29) in the desired optimum point when  $N \rightarrow +\infty$ . The result is summarized by the following Proposition.

*Proposition 1:* Let  $\mathbf{z}$  be distributed according to  $\mathcal{CN}(\mathbf{0}, \mathbf{C}_{\mathbf{z}})$ . Let  $f(\cdot)$  be a normalized model pdf so that  $\phi(\cdot) \leftarrow \nu^{-1} \phi(\cdot)$ ,  $\mathbf{w}$  be the true separating vector such that  $s = \mathbf{w}^H \mathbf{x}$ , and  $N \rightarrow +\infty$ . Then, the Hessian matrices of (19) defined as  $\mathbf{H}_1 = \frac{\partial^2 \mathcal{C}_{\text{OG}}}{\partial \mathbf{w}^T \partial \mathbf{w}}$  and  $\mathbf{H}_2 = \frac{\partial^2 \mathcal{C}_{\text{OG}}}{\partial \mathbf{w}^H \partial \mathbf{w}}$  are equal to

$$\mathbf{H}_1 = (c_3 \mathbf{a} \mathbf{a}^T)^*, \quad (30)$$

$$\mathbf{H}_2 = (c_1 \mathbf{C} + c_2 \mathbf{a} \mathbf{a}^H)^T, \quad (31)$$

where

$$c_1 = \frac{1}{\sigma^2} \left( \frac{\nu - \rho}{\nu} \right), \quad (32)$$

$$c_2 = -\sigma^2 c_1 - c_3, \quad (33)$$

$$c_3 = \frac{1}{2\nu} (\xi - \eta - \nu), \quad (34)$$

and

$$\rho = \mathbb{E} \left[ \frac{\partial \phi \left( \frac{s}{\sigma} \right)}{\partial s^*} \right], \quad (35)$$

$$\xi = \mathbb{E} \left[ \frac{\partial \phi \left( \frac{s}{\sigma} \right) |s|^2}{\partial s^* \sigma^2} \right], \quad (36)$$

$$\eta = \mathbb{E} \left[ \frac{\partial \phi \left( \frac{s}{\sigma} \right) s^2}{\partial s \sigma^2} \right]. \quad (37)$$

*Proof:* See Appendix A.

The proposed one-unit algorithm iterates in the direction inspired by the Newton-Raphson update [42]

$$\mathbf{w}_{\text{new}} = \mathbf{w} - \hat{\mathbf{H}}^{-1} (\nabla - \hat{\mathbf{H}}_1^* \hat{\mathbf{H}}_2^{-1} \nabla^*), \quad (38)$$

where  $\hat{\mathbf{H}} = \hat{\mathbf{H}}_2^* - \hat{\mathbf{H}}_1^* \hat{\mathbf{H}}_2^{-1} \hat{\mathbf{H}}_1$ ,  $\nabla$  is given by (29), and  $\hat{\mathbf{H}}_1$  and  $\hat{\mathbf{H}}_2$  are computed using the expressions (30) and (31), respectively, where (35)–(37) are replaced by their sample-based estimates. That means that the algorithm is not exactly the Newton-Raphson one, because the Hessian matrix is replaced by its analytic expression as if the current  $\mathbf{w}$  was the true separating vector.

In Appendix B, it is shown that

$$\hat{\mathbf{H}} = \left( \frac{\hat{\nu} - \hat{\rho}}{\hat{\nu}} \right)^* \left( \frac{\hat{\mathbf{C}}}{\hat{\sigma}^2} - \mathbf{a} \mathbf{a}^H \right), \quad (39)$$

and  $\hat{\mathbf{H}}_1^* \hat{\mathbf{H}}_2^{-1} \nabla^* = \mathbf{0}$ , so (38) is simplified to

$$\mathbf{w}_{\text{new}} = \mathbf{w} - \hat{\mathbf{H}}^{-1} \nabla. \quad (40)$$

However, the reader can notice that  $\hat{\mathbf{H}}$  is rank deficient, so  $\hat{\mathbf{H}}^{-1}$  actually does not exist. Indeed, for any value of  $\mathbf{w}$  (and  $\mathbf{a}$  linked through the OGC), the observed signals are equal to  $\mathbf{x} = \mathbf{a}\hat{s} + \hat{\mathbf{y}}$  where  $\hat{\mathbf{y}} = \mathbf{Q}\hat{\mathbf{z}}$ . The OGC guarantees that  $\hat{\mathbb{E}}[\hat{s}\hat{\mathbf{y}}] = 0$ , therefore,  $\hat{\mathbf{C}} = \hat{\sigma}^2\mathbf{a}\mathbf{a}^H + \hat{\mathbf{C}}_{\mathbf{y}}$ , where  $\hat{\mathbf{C}}_{\mathbf{y}} = \hat{\mathbb{E}}[\hat{\mathbf{y}}\hat{\mathbf{y}}^H]$ . So finally  $\hat{\mathbf{H}} \propto \hat{\mathbf{C}}_{\mathbf{y}}$ , whose rank is  $d - 1$ . This rank deficiency is caused by the scaling ambiguity of  $\mathbf{w}$ : There is a free scalar parameter with respect to which the contrast function is invariant. Fortunately, it appears that  $\nabla$  belongs to the column-space of  $\hat{\mathbf{H}}$ . After some algebra, we receive the following Proposition.

*Proposition 2:* The update (40) can be re-written as

$$\mathbf{w}_{\text{new}} = \mathbf{w} - \left( \frac{\hat{\nu}}{\hat{\nu} - \hat{\rho}} \right)^* \hat{\sigma}^2 \hat{\mathbf{C}}^{-1} \nabla. \quad (41)$$

*Proof:* See Appendix B.

Now, we get back to  $T \geq 1$  and  $K \geq 1$ . By inspecting (12), we can see that all terms with different  $t$  values are decoupled. The decoupling also holds for the dataset index  $k$  up to the first term in (12). However, since there is no coupling between the arguments of  $f(\{\cdot\}_k)$ , we only need to generalize the definition (25) to

$$\phi_k(\{s^{k,t}\}_k) = -\frac{\partial}{\partial s_k^*} \log f(\{s^{k,t}\}_k), \quad (42)$$

and, then, write all the other model parameters and signals' statistics with the superscript  $k, t$ . The gradient of (19) and the counterpart of the second-order derivative matrix (39) are equal to

$$\nabla^k = \left\langle \mathbf{a}^{k,t} - \frac{1}{\hat{\nu}^{k,t}} \hat{\mathbb{E}} \left[ \phi_k \left( \left\{ \frac{\hat{s}^{k,t}}{\hat{\sigma}^{k,t}} \right\}_k \right) \frac{\mathbf{x}^{k,t}}{\hat{\sigma}^{k,t}} \right] \right\rangle_t, \quad (43)$$

$$\hat{\mathbf{H}}^k = \left\langle \left( \frac{\hat{\nu}^{k,t} - \hat{\rho}^{k,t}}{\hat{\nu}^{k,t}} \right)^* \left( \frac{\hat{\mathbf{C}}^{k,t}}{(\hat{\sigma}^{k,t})^2} - \mathbf{a}^{k,t} (\mathbf{a}^{k,t})^H \right) \right\rangle_t. \quad (44)$$

Similar to (39), the scaling ambiguity causes the rank of (44) to be exactly equal to  $d - 1$ . However, we can follow the same approach as the one used in Proposition 2 to justify that the update for  $T \geq 1$  and  $K \geq 1$  is

$$\mathbf{w}_{\text{new}}^k = \mathbf{w}^k - \left\langle \left( \frac{\hat{\nu}^{k,t} - \hat{\rho}^{k,t}}{\hat{\nu}^{k,t}} \right)^* \frac{\hat{\mathbf{C}}^{k,t}}{(\hat{\sigma}^{k,t})^2} \right\rangle_t^{-1} \nabla^k. \quad (45)$$

Given the initial value of  $\mathbf{w}^k$ , for all  $k = 1, \dots, K$ , the proposed algorithm proceeds by computing (14),  $\hat{s}^{k,t} = (\mathbf{w}^k)^H \mathbf{x}^{k,t}$ ,  $\hat{\sigma}^{k,t}$  by (11),  $\hat{\nu}^{k,t}$  and  $\hat{\rho}^{k,t}$  according to (24) and (35), respectively, and updates the separating vectors through (43) and (45). The separating vectors can be normalized so that, for example, the scale of the SOI over all blocks equals one. The updates are repeated until the stopping rule from [26] is satisfied for all  $k = 1, \dots, K$ . The algorithm is referred to as one-unit FastDIVA.

### F. Relationship to One-Unit FastICA/FastIVA

One-unit FastICA is designed for BSE for the case  $K = 1$  and  $T = 1$  (the indices  $k$  and  $t$  can be omitted here). When the input signals have been pre-whitened so that  $\hat{\mathbf{C}} = \mathbf{I}_d$  [5], the one-unit

FastICA update rule is

$$\mathbf{w}_{\text{new}} = \hat{\mathbb{E}}[\phi(\hat{s})\mathbf{x}] - \rho\mathbf{w} \quad (46)$$

for the real-valued case [26], and

$$\mathbf{w}_{\text{new}} = \hat{\mathbb{E}}[\mathbf{x}g(|\hat{s}|^2)] - \hat{\mathbb{E}}[g(|\hat{s}|^2) + |\hat{s}|^2 g'(|\hat{s}|^2)]\mathbf{w} \quad (47)$$

for the complex-valued case [44], where  $g(\cdot)$  is the derivative of the contrast function, which is a real-valued smooth even function of  $|\hat{s}|^2$ . After each update,  $\mathbf{w}$  is normalized, which is equivalent to  $\hat{\sigma} = 1$  since  $\hat{\mathbf{C}} = \mathbf{I}$ .

We can compare (45) in a similar setting when  $\hat{\sigma} = 1$  and  $\hat{\mathbf{C}} = \mathbf{I}_d$ . The OGC (14) is then translated to  $\mathbf{a} = \mathbf{w}$ , and (45) is simplified to

$$\mathbf{w}_{\text{new}} = \mathbf{w} - \left( \frac{\hat{\nu}}{\hat{\nu} - \hat{\rho}} \right)^* (\mathbf{w} - \hat{\nu}^{-1} \hat{\mathbb{E}}[\phi(\hat{s})\mathbf{x}]), \quad (48)$$

Since the scale of  $\mathbf{w}_{\text{new}}$  can be arbitrary (the vector can be normalized afterwards), the right-hand side of (48) can be multiplied by the scalar factor  $(\hat{\nu} - \hat{\rho})^*$ , which, after a few simplifications, results in

$$\mathbf{w}_{\text{new}} = \hat{\mathbb{E}}[\phi(\hat{s})\mathbf{x}] - \rho^*\mathbf{w}. \quad (49)$$

It is worth noting here that  $\rho$  should be real-valued, provided that the model density  $f(\cdot)$  is a real-valued function. Once  $f(\cdot) = f(\cdot)^*$ , it holds that  $\rho = \rho^*$  [45]. By comparing (49) with (46), we can see that the update rules of one-unit FastDIVA and one-unit FastICA are the same in the real-valued case.

The complex-valued FastICA was derived in a different way, assuming a constrained class of contrast functions suitable for circular sources. The update rule (47) is different from (49). The latter is actually simpler and valid for circular as well as non-circular SOI (and a circular background).

For  $T = 1$  and  $K \geq 1$ , similar conclusions hold when comparing the update rules of FastIVA derived in [27] (Equation 58 in [27]), which are similar to (47), while (45) is simplified (when  $T = 1$ ,  $\hat{\mathbf{C}}^k = \mathbf{I}_d$  and  $\sigma^k = 1$ ) to

$$\mathbf{w}_{\text{new}}^k = \hat{\mathbb{E}}[\phi_k(\{\hat{s}^k\}_k)\mathbf{x}^k] - (\rho^k)^*\mathbf{w}^k. \quad (50)$$

To conclude, one-unit FastDIVA is an extension of FastICA and FastIVA for  $T > 1$  under the CSV model, in the real-valued case, and an extension and simplification involving non-circular SOI, in the complex-valued case.

### G. Separation of Several Signals

We now focus on the BSS problem when  $1 \leq r \leq d$  independent signals should be separated from each other and from the remainder of the signal (i.e., the other components and the noise). Following the idea of [26], [46], we propose to run  $r$  one-unit algorithms successively or in parallel while preventing them from extracting the same sources. To this end, the orthogonality constraint is imposed [47].

Throughout this Subsection, we will omit the dataset index  $k$  as the proposed approaches operate independently in each dataset.

1) *Symmetric Approach*: The approach presented here is suitable for dynamic mixtures (2) satisfying condition (C1) as defined in Section II. The deflation and symmetric approaches can then be used as they were designed for the static case  $T = 1$  [26]. Let us recall the symmetric approach here (the deflation approach can be derived similarly [26]).

Consider  $r$  separating vectors  $\mathbf{w}_1, \dots, \mathbf{w}_r$  each being updated through (45). Since the output signals, denoted as  $\hat{s}_1^t = \mathbf{w}_1^H \mathbf{x}^t, \dots, \hat{s}_r^t = \mathbf{w}_r^H \mathbf{x}^t$ , should be independent, it is reasonable that their mutual correlations estimated over all available samples (and blocks) should be constrained to equal zero. Specifically, the condition is that

$$\langle \hat{\mathbf{E}}[\hat{s}_i^t (\hat{s}_j^t)^*] \rangle_t = \langle \mathbf{w}_i^H \hat{\mathbf{C}}^t \mathbf{w}_j \rangle_t = \mathbf{w}_i^H \mathbf{R} \mathbf{w}_j = \delta_{ij}, \quad (51)$$

where  $\mathbf{R} = \langle \hat{\mathbf{C}}^t \rangle_t$  and  $\delta_{ij}$  denotes the Kronecker symbol, and  $i, j = 1, \dots, r$ . Let  $\mathbf{W}^+ = [\mathbf{w}_1, \dots, \mathbf{w}_r]$  involve the separating vectors after they were updated through (45), which do not satisfy (51), in general. The symmetric approach therefore proceeds by

$$\mathbf{W}_{\text{new}} = \mathbf{W}^+ ((\mathbf{W}^+)^H \mathbf{R} \mathbf{W}^+)^{-\frac{1}{2}}. \quad (52)$$

Since  $\mathbf{W}_{\text{new}}^H \mathbf{R} \mathbf{W}_{\text{new}} = \mathbf{I}_r$ , the columns of  $\mathbf{W}_{\text{new}}$  satisfy (51) and, therefore, can be used as the orthogonalized counterparts of  $\mathbf{W}^+$ .

Symmetric FastDIVA, as the proposed method to separate  $r$  independent signals is called, alternates between the updates of the separating vectors according to (50) and their subsequent orthogonalizations (52), until convergence.

2) *Block-Deflation Approach*: This approach is tailored to mixtures (2) satisfying condition (C2) as defined in Section II. It imposes a stronger condition on the extracted signals by making them orthogonal separately in each block. Specifically, it is expected that

$$\hat{\mathbf{E}}[\hat{s}_i^t (\hat{s}_j^t)^*] = \delta_{ij} (\hat{\sigma}_i^{k,t})^2 \quad (53)$$

for every  $t = 1, \dots, T$  and  $i, j = 1, \dots, r$ .

To this end, we propose an extended, so-called, *block deflation* scheme, which proceeds as follows. The first signal is extracted from the original data by one-unit FastDIVA. The extracted signal is then subtracted from the original input signals (on each block) using least-squares projections. Then, one-unit FastDIVA is applied to the new data and extracts the second signal, whose orthogonality is ensured due to the projection properties. This process is repeated recursively until  $r$  signals are extracted.

Let  $\mathbf{x}_i^t$  denote the input signals on the  $t$ th block at the  $i$ th stage of the block-deflation scheme, and let  $\mathbf{w}_i$  be the separating vector obtained after one iteration by one-unit FastDIVA applied to  $\mathbf{x}_i^t$ . For  $i = 1$ ,  $\mathbf{x}_i^t = \mathbf{x}^t$  (the original input data). The new data  $\mathbf{x}_{i+1}^t$  are obtained by the least-squares subtraction of  $\hat{s}_i^t = \mathbf{w}_i^H \mathbf{x}_i^t$  from  $\mathbf{x}_i^t$ . Owing to the OGC (14) imposed between the mixing and separating vectors of the extracted source, the new data is obtained through

$$\mathbf{x}_{i+1}^t = \mathbf{\Pi}_i^t \mathbf{x}_i^t, \quad (54)$$

where  $\mathbf{\Pi}_i^t = \mathbf{E}_i(\mathbf{I}_{d-i+1} - \hat{\mathbf{a}}_i^t \mathbf{w}_i^H)$ ;  $\hat{\mathbf{a}}_i^t$  is the estimated mixing vector on the  $t$ th block corresponding to the  $i$ th extracted signal

with respect to data  $\mathbf{x}_i^t$ .  $\mathbf{E}_i$  is a suitable  $(d-i) \times (d-i+1)$  matrix having the full row-rank; it reduces the dimension of  $\mathbf{x}_{i+1}^t$  as compared to  $\mathbf{x}_i^t$  by one (so that the new data is not rank deficient); the dimension of  $\mathbf{x}_i^t$  is  $d-i+1$ .

The estimated vectors  $\mathbf{w}_i$  and  $\hat{\mathbf{a}}_i^t$  operate on the data  $\mathbf{x}_i^t$ . In order to derive their counterparts operating on the original data  $\mathbf{x}^t$ , let us introduce the following definitions:

$$\mathbf{P}_1^t = \mathbf{I}_d, \quad (55)$$

$$\mathbf{P}_i^t = \mathbf{\Pi}_{i-1}^t \mathbf{\Pi}_{i-2}^t \dots \mathbf{\Pi}_1^t, \quad i > 1 \quad (56)$$

$$\hat{\mathbf{C}}_i^t = \hat{\mathbf{E}}[\mathbf{x}_i^t (\mathbf{x}_i^t)^H], \quad (57)$$

$$\mathbf{w}_i^t = (\mathbf{P}_i^t)^H \mathbf{w}_i, \quad (58)$$

$$\mathbf{a}_i^t = \frac{\hat{\mathbf{C}}_i^t \mathbf{w}_i^t}{(\mathbf{w}_i^t)^H \hat{\mathbf{C}}_i^t \mathbf{w}_i^t}. \quad (59)$$

It is then straightforward to verify that, for  $i = 1, \dots, r$ ,

$$\mathbf{x}_i^t = \mathbf{P}_i^t \mathbf{x}^t, \quad (60)$$

$$\hat{s}_i^t = \mathbf{w}_i^H \mathbf{x}_i^t = (\mathbf{w}_i^t)^H \mathbf{x}^t, \quad (61)$$

$$\hat{\mathbf{C}}_i^t = \mathbf{P}_i^t \hat{\mathbf{C}}^t (\mathbf{P}_i^t)^H, \quad (62)$$

$$\hat{\mathbf{a}}_i^t = \mathbf{P}_i^t \mathbf{a}_i^t. \quad (63)$$

Note that  $\mathbf{w}_i$  and  $\hat{\mathbf{a}}_i^t$  operate on  $\mathbf{x}_i^t$  and, since both have been estimated by one-unit FastDIVA, they are coupled through the OGC, i.e.,  $\hat{\mathbf{a}}_i^t = \frac{\hat{\mathbf{C}}_i^t \mathbf{w}_i}{(\mathbf{w}_i)^H \hat{\mathbf{C}}_i^t \mathbf{w}_i}$ . In addition,  $\mathbf{w}_i$  is independent of  $t$  due to the CSV model assumed by one-unit FastDIVA.

The counterpart of  $\mathbf{w}_i$  and  $\hat{\mathbf{a}}_i^t$  operating on  $\mathbf{x}^t$  is  $\mathbf{w}_i^t$  and  $\mathbf{a}_i^t$ , respectively. Interestingly, unless  $i = 1$  holds,  $\mathbf{w}_i^t$  is, in general, no longer independent of  $t$ .

#### IV. PERFORMANCE ANALYSIS

The goal here is to analyze the accuracy of one-unit FastDIVA considering the BSE problem under the CSV mixture model. The accuracy is studied by analyzing the mean residual presence of the  $j$ th original signal in the extracted signal  $j = 1, \dots, d$ , which is characterized by the mean interference-to-signal ratio (ISR) achieved by the algorithm.

To this end, we compute the asymptotic variance of the estimated separating vector that is obtained by the algorithm as the optimum point of the contrast function (19); it is assumed that  $N \rightarrow \infty$ , which means, for a fixed value of  $T$ , that also  $N_b \rightarrow \infty$ . Using the equivariance property of the BSE problem, proven in [25], we consider the special case as if the true mixing and separating vectors were  $\mathbf{a}^{k,t} = \mathbf{w}^k = [1; \mathbf{0}]$ ,  $k = 1, \dots, K$  (in Section V, this analysis is verified for general mixing and separating vectors). Then, by (3) it follows that  $\mathbf{x}^{k,t} = [s^{k,t}; -\mathbf{z}^{k,t}]$ .

Let  $\hat{\mathbf{w}}^k, \hat{s}^{k,t} = (\hat{\mathbf{w}}^k)^H \mathbf{x}^{k,t}, \hat{\sigma}_s^{k,t}$  and  $\hat{\mathbf{z}}^{k,t}$  denote, respectively, the estimates of  $\mathbf{w}^k, s^{k,t}$ , of the sample-based variance estimate of  $\hat{s}^{k,t}$ , and of the background signals. The following notation will be used:

$$\hat{\nu}_s^{k,t} = \hat{\mathbf{E}} \left[ \phi_k \left( \left\{ \frac{\hat{s}^{k,t}}{\hat{\sigma}_s^{k,t}} \right\}_k \right) \frac{\hat{s}^{k,t}}{\hat{\sigma}_s^{k,t}} \right], \quad (64)$$

$$\hat{\rho}_s^{k,t} = \hat{\mathbb{E}} \left[ \frac{\partial \phi_k}{\partial s_k^*} \left( \left\{ \frac{\hat{s}^{k,t}}{\hat{\sigma}_s^{k,t}} \right\}_k \right) \right]. \quad (65)$$

Next, we introduce the random variables derived from the samples of  $s^{k,t}$  and  $\mathbf{z}^{k,t}$

$$\hat{\chi}^{k,t} = \hat{\mathbb{E}} [\mathbf{z}^{k,t} (s^{k,t})^H] \quad (66)$$

$$\hat{\zeta}^{k,t} = \hat{\mathbb{E}} \left[ \phi_k \left( \left\{ \frac{s^{k,t}}{\sigma^{k,t}} \right\}_k \right) \frac{\mathbf{z}^{k,t}}{\sigma^{k,t}} \right] \quad (67)$$

$$\hat{\zeta}_s^{k,t} = \hat{\mathbb{E}} \left[ \phi_k \left( \left\{ \frac{\hat{s}^{k,t}}{\hat{\sigma}_s^{k,t}} \right\}_k \right) \frac{\mathbf{z}^{k,t}}{\hat{\sigma}_s^{k,t}} \right]. \quad (68)$$

Let the structure of  $\hat{\mathbf{w}}^k$  be

$$\hat{\mathbf{w}}^k = [1 + p^k, \mathbf{q}^k]^T \quad (69)$$

where  $p^k$  and  $\mathbf{q}^k$  are random variables of the stochastic order  $O_p(N_b^{-1})$  and  $O_p(N_b^{-1/2})$ , respectively;  $O_p(\cdot)$  represents the stochastic order symbol; see Appendix C in [48]. The goal now is to express  $p^k$  and  $\mathbf{q}^k$  as functions of  $s^{k,t}$  and  $\mathbf{z}^{k,t}$  and to compute their asymptotic variances. Finally, only the asymptotic covariance of  $\mathbf{q}^k$  will be needed.

Note that  $\hat{\chi}^{k,t}$  and  $\hat{\zeta}^{k,t}$  have the same stochastic order below. We can write

$$\hat{\mathbf{C}}^{k,t} = \hat{\mathbb{E}}[\mathbf{x}^{k,t} (\mathbf{x}^{k,t})^H] = \begin{bmatrix} (\sigma^{k,t})^2 + c^{k,t} & -(\hat{\chi}^{k,t})^H \\ -\hat{\chi}^{k,t} & \mathbf{C}_z^{k,t} + \Xi^{k,t} \end{bmatrix}, \quad (70)$$

where

$$c^{k,t} = \hat{\mathbb{E}}[s^{k,t} (s^{k,t})^H] - (\sigma^{k,t})^2, \quad (71)$$

$$\Xi^{k,t} = \hat{\mathbb{E}}[\mathbf{z}^{k,t} (\mathbf{z}^{k,t})^H] - \mathbf{C}_z^{k,t}. \quad (72)$$

Define the difference between the sample-based variances as

$$b^{k,t} = \hat{\sigma}^{k,t} - \hat{\sigma}_s^{k,t} = \hat{\mathbb{E}}[s^{k,t} (s^{k,t})^H] - \hat{\mathbb{E}}[\hat{s}^{k,t} (\hat{s}^{k,t})^H] \\ = \hat{\mathbb{E}}[(\mathbf{q}^k)^H \mathbf{z}^{k,t} (\mathbf{z}^{k,t})^H \mathbf{q}^k]. \quad (73)$$

The stochastic order of  $c^{k,t}$  and  $\Xi^{k,t}$  is the same as that of  $\hat{\chi}^{k,t}$ , i.e.,  $O_p(N_b^{-1/2})$ , while  $b^{k,t}$  is of order  $O_p(N_b^{-1})$ . It holds that

$$\frac{\hat{s}^{k,t}}{\hat{\sigma}_s^{k,t}} = \frac{(\hat{\mathbf{w}}^k)^H \mathbf{x}^{k,t}}{\sigma^{k,t} + b^{k,t} + c^{k,t}} = \frac{s^{k,t}}{\sigma^{k,t}} - \frac{(\mathbf{q}^k)^H \mathbf{z}^{k,t}}{\sigma^{k,t}} + o_p(N_b^{-1/2}). \quad (74)$$

Assuming the smoothness of  $\phi$  and using the first-order Taylor series expansion, we get

$$\phi_k \left( \left\{ \frac{\hat{s}^{k,t}}{\hat{\sigma}_s^{k,t}} \right\}_k \right) = \phi_k \left( \left\{ \frac{s^{k,t}}{\sigma^{k,t}} \right\}_k \right) \\ - \frac{1}{\sigma^{k,t}} \sum_{k=1}^K (\mathbf{q}^k)^H \mathbf{z}^{k,t} \frac{\partial \phi_k}{\partial s_k} \left( \left\{ \frac{s^{k,t}}{\sigma^{k,t}} \right\}_k \right) \\ - \frac{1}{\sigma^{k,t}} \sum_{k=1}^K (\mathbf{z}^{k,t})^H \mathbf{q}^k \frac{\partial \phi_k^*}{\partial s_k} \left( \left\{ \frac{s^{k,t}}{\sigma^{k,t}} \right\}_k \right) + o_p(N_b^{-1/2}). \quad (75)$$

From the uncorrelatedness of datasets, and assuming the circularity of  $\mathbf{z}^{k,t}$ , i.e  $\mathbb{E}[\mathbf{z}^{k,t} (\mathbf{z}^{k,t})^T] = \mathbf{0}$ , we can write

$$\hat{\zeta}_s^{k,t} = \hat{\mathbb{E}} \left[ \phi_k \left( \left\{ \frac{\hat{s}^{k,t}}{\hat{\sigma}_s^{k,t}} \right\}_k \right) \frac{\mathbf{z}^{k,t}}{\hat{\sigma}_s^{k,t}} \right] \\ = \hat{\mathbb{E}} \left[ \phi_k \left( \left\{ \frac{s^{k,t}}{\sigma^{k,t}} \right\}_k \right) \frac{\mathbf{z}^{k,t}}{\sigma^{k,t}} \right] \\ - \frac{1}{(\sigma^{k,t})^2} \hat{\mathbb{E}} \left[ \frac{\partial \phi_k}{\partial s_k} \left( \left\{ \frac{s^{k,t}}{\sigma^{k,t}} \right\}_k \right) \mathbf{z}^{k,t} (\mathbf{z}^{k,t})^T (\mathbf{q}^k)^* \right] \\ - \frac{1}{(\sigma^{k,t})^2} \hat{\mathbb{E}} \left[ \frac{\partial \phi_k}{\partial s_k^*} \left( \left\{ \frac{s^{k,t}}{\sigma^{k,t}} \right\}_k \right) \mathbf{z}^{k,t} (\mathbf{z}^{k,t})^H \mathbf{q}^k \right] + o_p(N_b^{-1/2}) \\ = \hat{\zeta}^{k,t} - \frac{\hat{\rho}^{k,t} \mathbf{C}_z^{k,t} \mathbf{q}^k}{(\sigma^{k,t})^2} + o_p(N_b^{-1/2}). \quad (76)$$

Next,

$$\hat{\mathbf{C}}^{k,t} \mathbf{w}^k = \begin{bmatrix} (\sigma^{k,t})^2 + c^{k,t} \\ -\hat{\chi}^{k,t} + \mathbf{C}_z^{k,t} \mathbf{q}^k \end{bmatrix} + o_p(N_b^{-1/2}) \quad (77)$$

$$(\mathbf{w}^k)^H \hat{\mathbf{C}}^{k,t} \mathbf{w}^k = (\sigma^{k,t})^2 + c^{k,t} + o_p(N_b^{-1/2}). \quad (78)$$

The mixing vector estimated by (14) and the gradient (43) can be expressed, using (76)–(78), respectively, as

$$\hat{\mathbf{a}}^{k,t} = \frac{1}{(\sigma^{k,t})^2} \begin{bmatrix} (\sigma^{k,t})^2 \\ -\hat{\chi}^{k,t} + \mathbf{C}_z^{k,t} \mathbf{q}^k \end{bmatrix} + o_p(N_b^{-1/2}), \quad (79)$$

$$\hat{\mathbf{V}}^k = \left\langle \mathbf{a}^{k,t} - \begin{bmatrix} 1 \\ -\hat{\zeta}_s^{k,t} / \hat{\nu}_s^{k,t} \end{bmatrix} \right\rangle_t \\ = \left\langle \begin{bmatrix} 0 \\ -\hat{\chi}^{k,t} + \mathbf{C}_z^{k,t} \mathbf{q}^k \\ -\hat{\zeta}_s^{k,t} / \hat{\nu}_s^{k,t} \end{bmatrix} \right\rangle_t + o_p(N_b^{-1/2}). \quad (80)$$

The stationary point of the algorithm is now sought as the solution of  $\hat{\mathbf{V}}^k = \mathbf{0}$ . This gives us

$$\mathbf{q}^k = \mathbf{R}^{k,t} \left\langle \frac{\hat{\nu}_s^{k,t} \hat{\chi}^{k,t} - \hat{\zeta}_s^{k,t} (\sigma^{k,t})^2}{(\sigma^{k,t})^2 \hat{\nu}_s^{k,t}} \right\rangle_t + o_p(N_b^{-1/2}), \quad (81)$$

where

$$\mathbf{R}^{k,t} = \left\langle \mathbf{C}_z^{k,t} \frac{\nu^{k,t} - \rho^{k,t}}{(\sigma^{k,t})^2 \nu^{k,t}} \right\rangle_t^{-1}. \quad (82)$$

Computation of the asymptotic covariance of  $\mathbf{q}^k$  remains to be done. Straightforward computations give

$$\mathbb{E}[\hat{\chi}^{k,t} (\hat{\chi}^{k,t})^H] = \frac{1}{N_b} (\sigma^{k,t})^2 \mathbf{C}_z^{k,t} \quad (83)$$

$$\mathbb{E}[\hat{\zeta}^{k,t} (\hat{\zeta}^{k,t})^H] = \frac{1}{N_b} \frac{\varphi^{k,t}}{(\sigma^{k,t})^2} \mathbf{C}_z^{k,t} \quad (84)$$

$$\mathbb{E}[\hat{\chi}^{k,t} (\hat{\zeta}^{k,t})^H] = \frac{\nu^{k,t}}{N_b} \mathbf{C}_z^{k,t}, \quad (85)$$

where we have introduced one more statistic related to the SOI

$$\varphi^{k,t} = \mathbb{E} \left[ \left| \phi_k \left( \left\{ \frac{s^{k,t}}{\sigma_s^{k,t}} \right\}_k \right) \right|^2 \right]. \quad (86)$$

Using these expressions and (81), the asymptotic covariance of  $\mathbf{q}^k$  is given by

$$\begin{aligned} \text{cov}[\mathbf{q}^k] &= \text{cov} \left[ \mathbf{R}^{k,t} \left\langle \frac{\hat{\nu}^{k,t} \hat{\chi}^{k,t} - \hat{\zeta}^{k,t} (\sigma^{k,t})^2}{(\sigma^{k,t})^2 \hat{\nu}^{k,t}} \right\rangle_t \right] \\ &= \frac{1}{N} \mathbf{R}^{k,t} \left\langle \mathbf{C}_{\mathbf{z}}^{k,t} \frac{\varphi^{k,t} - |\nu^{k,t}|^2}{(\sigma^{k,t})^2 |\nu^{k,t}|^2} \right\rangle_t (\mathbf{R}^{k,t})^H + o(N^{-1}). \end{aligned} \quad (87)$$

The theoretical ISR reads

$$\begin{aligned} \text{ISR}^k &= \frac{\sum_{t=1}^T \text{E} \left[ |(\hat{\mathbf{w}}^k)^H \mathbf{y}^{k,t}|^2 \right]}{\sum_{t=1}^T \text{E} \left[ |(\hat{\mathbf{w}}^k)^H \mathbf{a}^{k,t} s_{k,t}|^2 \right]} \\ &= \frac{\sum_{t=1}^T \text{tr} \left[ \mathbf{C}_{\mathbf{z}}^{k,t} \text{cov}[\mathbf{q}^k] \right]}{\sum_{t=1}^T (\sigma^{k,t})^2} = \frac{\text{tr} \left[ \langle \mathbf{C}_{\mathbf{z}}^{k,t} \rangle_t \text{cov}[\mathbf{q}^k] \right]}{\langle (\sigma^{k,t})^2 \rangle_t}. \end{aligned} \quad (88)$$

Hence, using (82) and (87), the asymptotic mean ISR achieved by the algorithm is

$$\begin{aligned} \text{E} [\text{ISR}^k] &\approx \frac{1}{N} \text{tr} \left[ \frac{\langle \mathbf{C}_{\mathbf{z}}^{k,t} \rangle_t}{\langle (\sigma^{k,t})^2 \rangle_t} \left\langle \mathbf{C}_{\mathbf{z}}^{k,t} \frac{\nu^{k,t} - \rho^{k,t}}{(\sigma^{k,t})^2 \nu^{k,t}} \right\rangle_t^{-1} \right. \\ &\quad \left. \left\langle \mathbf{C}_{\mathbf{z}}^{k,t} \frac{\varphi^{k,t} - |\nu^{k,t}|^2}{(\sigma^{k,t})^2 |\nu^{k,t}|^2} \right\rangle_t \left( \left\langle \mathbf{C}_{\mathbf{z}}^{k,t} \frac{\nu^{k,t} - \rho^{k,t}}{(\sigma^{k,t})^2 \nu^{k,t}} \right\rangle_t^{-1} \right)^H \right]. \end{aligned} \quad (89)$$

To compare this result with previous analyses, consider  $T = K = 1$ . Then, (89) is simplified to

$$\text{E} [\text{ISR}^k] \approx \frac{d-1}{N} \frac{\varphi^k - |\nu^k|^2}{|\nu^k - \rho^k|^2}, \quad (90)$$

which coincides with the results given in [17], [49], [50] (that result is also confirmed in the present paper for the complex-valued case and for  $K > 1$ ).

Next, let the model density  $f(\cdot)$  correspond to the normalized true pdf of the SOI for all  $k$  and  $t$ ; let us, for the moment, denote this normalized true pdf by  $p^{k,t}(\cdot)$ . Then the equalities  $\nu^{k,t} = 1$ ,  $\rho^{k,t} = \kappa^{k,t}$  and  $\varphi^{k,t} = \kappa^{k,t}$  hold, where

$$\kappa^{k,t} = \text{E} \left[ \left| \frac{\partial \log p^{k,t}(\{s_k\}_k)}{\partial s_k^*} \right|^2 \right]. \quad (91)$$

Formula (89) now takes on the form

$$\text{E} [\text{ISR}^k] \approx \frac{1}{N} \text{tr} \left[ \frac{\langle \mathbf{C}_{\mathbf{z}}^{k,t} \rangle_t}{\langle (\sigma^{k,t})^2 \rangle_t} \left\langle \mathbf{C}_{\mathbf{z}}^{k,t} \frac{\kappa^{k,t} - 1}{(\sigma^{k,t})^2} \right\rangle_t^{-1} \right]. \quad (92)$$

For  $K = 1$ , (92) coincides with the Cramér-Rao Lower Bound derived in [25] (70 in [25]), which points to the asymptotic efficiency of one-unit FastDIVA under the corresponding statistical (and mixing) model when the used nonlinearity corresponds with the true normalized score function of the SOI.

## V. NUMERICAL VALIDATION

In experiments, we simulate BSE and BSS on mixtures obeying dynamic models discussed in this paper. In BSE, one-unit

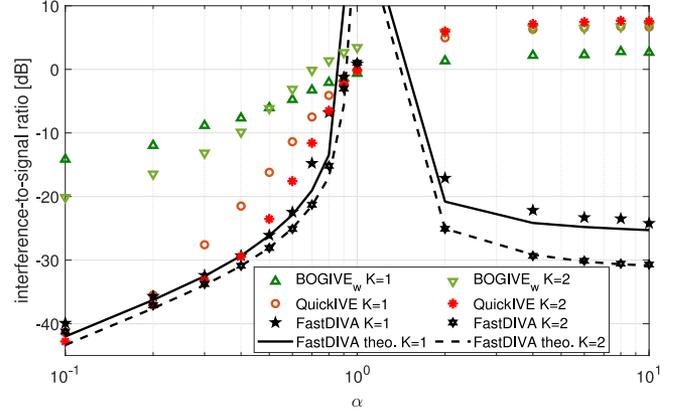


Fig. 4. Average ISR over 1000 as a function of  $\alpha$ , the shape parameter of the pdf of the SOI;  $\alpha = 1$  corresponds to Gaussian SOI, which is not identifiable. The pdf super-Gaussian and sub-Gaussian for  $\alpha < 1$  and  $\alpha > 1$ , respectively. “FastDIVA theo.” stands for the analytical prediction (89).

FastDIVA is compared with recent methods assuming CSV mixing, namely, with the gradient-based BOGIVE<sub>w</sub> [23] and with a more advanced QuickIVE [51]. FastICA/FastIVA are compared in the BSS tasks, both implemented as FastDIVA with a special setting (i.e., when  $T = 1$  is assumed). All the algorithms use the rational nonlinearity given by [52]

$$\phi_k(\{s_k\}_k) = \frac{s_k^*}{1 + \sum_{k=1}^K |s_k|^2}. \quad (93)$$

The number of iterations is restricted to 100 in QuickIVE and in FastDIVA and to 1000 in BOGIVE<sub>w</sub>. The step size in BOGIVE<sub>w</sub> is set to 0.1.

The accuracy of separated signals is evaluated, after resolving the unknown order, in terms of ISR, as defined by the first fraction in (88) (the expectations are replaced by sample averages).

### A. Dynamic Blind Source Extraction

The simulation here is focused on the BSE problem to verify the efficiency of one-unit FastDIVA, to verify its analysis provided in Section IV, and to evaluate its speed. In a trial, a mixture of dimension  $d = 6$  is generated such that it obeys CSV with  $T = 5$  blocks of length  $N_b = 2000$ , i.e.,  $N = 10^4$ . The background signals are circular Gaussian while the SOI is generated according to the complex-valued Generalized Gaussian distribution [53] with the shape parameter  $\alpha$ , denoted as GG( $\alpha$ ). The variance of SOI is block-dependent, namely, equal to  $|\cos(i/6 * \pi)| + 1 - \sqrt{3}/2$  on the  $i$ th block. The mixing matrices are randomly generated so that the first rows of their inverse matrices are the same in all blocks, that is, (7) is satisfied.

The experiment is realized in two variants with  $K = 1$  and  $K = 2$ . In the latter case, the SOIs are, in both mixtures, rotated by a random unitary matrix (before they are mixed with the background) in order to establish their higher-order dependence. The compared methods are initialized by randomly perturbed true separating vectors, where the elements of the perturbations are  $\mathcal{CN}(0, 0.1)$ .

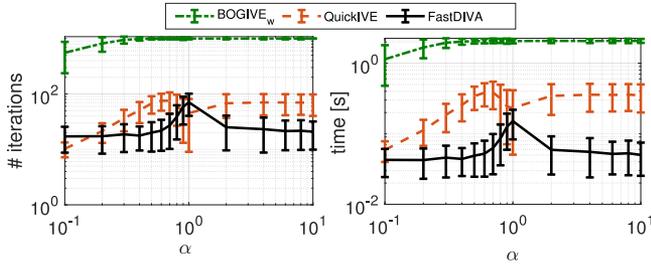


Fig. 5. The number of iterations and computational time needed by the compared methods to achieve convergence. Note that the maximum number of iteration is 100 for FastDIVA and QuickIVE and 1000 for BOGIVE<sub>w</sub>; simulations were done in Matlab R2020a on a server with Inter Xeon 12-core 2.6 GHz CPU, 64 GB RAM.

Fig. 4 shows ISR averaged<sup>9</sup> over 1000 trials as a function of  $\alpha \in [0.1, 10]$ . Note that the SOI is super-Gaussian for  $\alpha < 1$ , Gaussian for  $\alpha = 1$ , and sub-Gaussian for  $\alpha > 1$ . For  $\alpha = 1$ , the SOI is not identifiable. The average ISR established by the methods therefore tends to be close to or above 0 dB when  $\alpha$  is close to one, which means a poor extraction accuracy.

One-unit FastDIVA yields performance that is in good agreement with the theoretical analysis given by (89). BOGIVE<sub>w</sub> gives poor ISR compared to the other methods, because 1000 iterations is generally not sufficient to achieve the optimum point. QuickIVE achieves results similar to FastDIVA for  $\alpha < 0.3$  and slightly worse for  $\alpha \in [0.3, 1]$  (also because of the limited number of iterations). For  $\alpha > 1$ , BOGIVE<sub>w</sub> and QuickIVE fail to extract the SOI since the algorithms are not stable with respect to the SOI sub-Gaussianity and the nonlinearity (93). Here, FastDIVA inherits the stability of FastICA and works well also for  $\alpha > 1$ .

For  $K = 2$ , all methods achieve improved ISR as compared to the case of  $K = 1$ , which confirms the advantage following from the joint source extraction [54]. Fig. 5 shows the computational complexity in terms of the number of iterations and computational time. FastDIVA and QuickIVE show significantly faster convergence as compared to BOGIVE<sub>w</sub>, and FastDIVA is faster than QuickIVE.

### B. Dynamic Separation of Several Sources

Now, we focus on the BSS problem of  $1 \leq r \leq d$  signals from mixtures of dimension  $d$  obeying condition (C1) or (C2), as defined in Section II. In this scenario,  $d = 5$ ,  $T = 5$ ,  $K = 1$ ,  $N_b = 10^4$ ,  $N = 5 \cdot 10^4$ .

As for (C1),  $r$  complex-valued signals are generated according to GG(0.1) with the same variance profiles as the SOI in the previous experiment. The background is considered in two variants: Gaussian or GG(0.1). The mixing matrices are randomly generated so that the first  $r$  rows of their inverse matrices are the same in all blocks. For  $r = d$ , the entire matrices are the same in all blocks, which corresponds to the static mixing model.

<sup>9</sup>One percent of minimum and maximum values of ISR were discarded in order to eliminate the bias caused by the ambiguity of order (the algorithm might, in a few trials, be attracted by a different extreme of the contrast function corresponding to a signal different from the SOI).

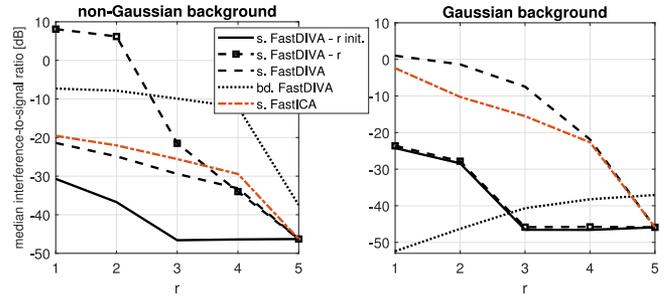


Fig. 6. Median ISR of 100 trials as a function of  $r$ ,  $r = 1, \dots, d$ ,  $d = 5$ , achieved by separating dynamic mixtures obeying condition (C1). For  $r = d = 5$ , the mixtures are static; “s.” and “bd.” stand for symmetric and block-deflation, respectively; “r” means that only  $r$  signals are being separated; “init” means a controlled initialization.

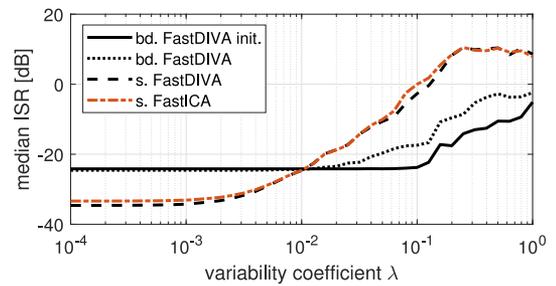


Fig. 7. Median ISR as a function of variability coefficient achieved in separation of real-valued mixtures of speech signals from Fig. 1 obeying condition (C2).

In the case of (C2), real-valued mixtures of the speech signals from Fig. 1 are considered. The mixing matrices are generated as follows. In the beginning, mixing and separating vectors  $\hat{\mathbf{a}}_i$  and  $\mathbf{w}_i$  of dimension  $i$  are generated at random, where  $i = 1, \dots, d$  such that  $\mathbf{w}_i^H \hat{\mathbf{a}}_i = 1$ ; their values remain fixed during the rest of the simulations. Then, in a trial, their values are perturbed by random vectors of the same size whose elements are taken from  $\mathcal{N}(0, \lambda^2)$ . The mixing vector  $\hat{\mathbf{a}}_i$  is perturbed differently on each block, which simulates a random walk of the associated source;  $\lambda$  thus plays the role of a variability coefficient of the mixture. The rows of de-mixing matrices are then obtained successively by using (55), (56) and (58). The mixing matrices are obtained as the inverse matrices of the de-mixing ones for  $i = 1, \dots, d$ . These steps guarantee that the mixtures obey (C2); for  $\lambda = 0$ , there are no temporal changes in the mixing model parameters, so the mixtures are static.

The results of these experiments are shown in Figures 6 and 7 in terms of median ISR computed over 100 trials for each  $r$  and  $\lambda$ , respectively. The median is used instead of the average because, in dynamic settings, the algorithms can fail in many more trials than in the static case; the results indicate that such failures mainly depend on the initializations. In legends, “s.” and “bd.” are acronyms for the symmetric and block-deflation variants, respectively; “r” means that only  $r$  signals are being separated; “init” means that the algorithm is initialized in a vicinity of the correct solution.

Fig. 6 shows that, in both background settings, symmetric FastDIVA yields excellent ISR on the mixtures obeying (C1) provided that it is properly initialized and the true number of signals that obey CSV is known. Without a proper initialization, its performance is close to symmetric FastICA, i.e., when  $T = 1$  blocks are assumed. In the static case  $r = d$ , the symmetric algorithms achieve the same superior performance (median ISR about  $-45$  dB); it is worth pointing out that, when  $r = d$ , FastDIVA assumes the overestimated number of blocks ( $T = 5$ ), nevertheless, this phenomenon does not deteriorate its performance. Block-deflation FastDIVA performs well in the Gaussian background setting and achieves a lower median ISR when  $r = d$  as compared to the symmetric algorithms. The latter observation agrees with the results of previous theoretical analyses of the symmetric and deflation approaches [50] (the symmetric one is usually more accurate; the accuracy of the deflation one depends on the order in which the signals are being separated).

Fig. 7 shows results of the experiment with mixtures (C2), which is, in fact, suitable for block-deflation FastDIVA. This algorithm tends to yield a constant median ISR until  $\lambda \approx 10^{-1}$ . This is indicative of the fact that the algorithm's performance is equivariant, i.e., independent of the mixing parameters, as are the theoretical bounds (89) and Cramér-Rao bounds in [25]. For higher values of  $\lambda$ , the probability grows for the algorithm getting stuck in a local extreme; this tendency deteriorates the median ISR. Symmetric FastDIVA and FastICA yield similar median ISR outputs in this scenario. For very small  $\lambda$  values, i.e., when the mixture is almost static, they achieve a better ISR than the block-deflation variant, which agrees with the observation shown in Fig. 6 for  $r = d$ . With growing  $\lambda$ , the performance of symmetric FastDIVA drops down because the mixture does not meet the condition (C1).

### C. Semi-Online Blind Source Extraction

Here, the application of CSV and one-unit FastDIVA is presented in an online BSE problem where a SOI is being extracted sequentially batch-by-batch. The choice of batch length plays an important role in such processing. It affects the key features of the online system: extraction accuracy, adaptability, and susceptibility to the discontinuity problem.

The benefit of using CSV in online processing is that it allows for dynamics within the batch by setting  $T > 1$ . Moreover, with  $T > 1$ , the length of batch can be increased without reducing the time-resolution of the on-line approach. We verify this feature in a simulated example where online BSE with  $T = 1$  and  $T > 1$  are compared.

The data are generated as follows. In one trial, a random instantaneous real-valued mixture ( $K = 1$ ) of dimension  $d = 10$  involving one moving laplacean SOI and 8 static interfering laplacean sources is generated. The sources have zero mean and unit variance. The mixing vector related to the SOI is continuously changing in a linear manner so that its value at the  $n$ th sample,  $n = 1, \dots, N$ , is

$$\mathbf{a}_n = \left(1 - \frac{n-1}{N-1}\right) \mathbf{a}_1 + \left(\frac{n-1}{N-1}\right) \mathbf{a}_N, \quad (94)$$

TABLE I  
RESULTS OF THE SIMULATED ONLINE BSE IN TERMS OF SINR AND SDR [DB]  
AVERAGED OVER 100 TRIALS

Angle $\angle(\mathbf{a}_1, \mathbf{a}_N)$			0°		10°		30°	
Batch length	shift	#CSV blocks ( $T$ )	SINR	SDR	SINR	SDR	SINR	SDR
500	100	1	13.8	25.2	11.3	21.3	6.3	9.0
2500	500	1	21.2	32.8	8.0	11.4	7.5	3.6
5000	1000	1	24.5	36.4	8.2	6.5	13.0	2.4
500	100	5	13.9	25.9	12.8	23.5	12.9	20.0
2500	500	5	21.2	32.8	18.4	24.8	19.7	13.9
5000	1000	5	24.3	36.4	20.4	25.0	23.0	17.6

where  $\mathbf{a}_1$  and  $\mathbf{a}_N$  are random vectors with unit norm. The movement speed of the SOI is controlled through the angular distance between  $\mathbf{a}_1$  and  $\mathbf{a}_N$ . Gaussian additive noise with zero mean and 0.1 variance is added to the mixture. The total length of data is  $N = 60\,000$  samples.

The separating vector is initialized by the LCMP beamformer [40] steered in the directions given by  $\mathbf{a}_1$  and  $\mathbf{a}_N$ . The data are then processed batch-by-batch with shift of a specific length; a useful choice when  $T > 1$  is  $N_b = L/T$  where  $L$  corresponds to the length of batch. One one-unit FastDIVA iteration per batch is performed, initialized by the separating vector from the previous batch.

The extraction accuracy is evaluated in terms of Signal-to-Interference-plus-Noise Ratio (SINR) and Signal-to-Distortion Ratio (SDR) where the latter is defined as

$$\text{SDR} = \frac{\widehat{\mathbb{E}}[\tilde{s}^2]}{\min_{\alpha} \widehat{\mathbb{E}}[(\tilde{s} - \alpha s)^2]}, \quad (95)$$

where  $\tilde{s}$  is the SOI component within the extracted signal  $\hat{s}$ , and  $s$  is the true SOI. Table I shows the results averaged over 100 trials as they depend on the batch length, shift, and the angle between  $\mathbf{a}_1$  and  $\mathbf{a}_N$ , denoted as  $\angle(\mathbf{a}_1, \mathbf{a}_N)$ .

When  $\angle(\mathbf{a}_1, \mathbf{a}_N) = 0$ , the mixing vector remains constant over time, so the mixture is static. Here, the SINR and SDR are obviously increasing with the growing batch length. As expected,  $T = 5$  brings no advantage compared to  $T = 1$ , in this case. By contrast, when the SOI is moving and  $\angle(\mathbf{a}_1, \mathbf{a}_N) > 0$ , the processing with  $T = 5$  brings significantly better SINR as well as SDR compared to  $T = 1$ .

Also, by detailed inspection of the values of SDR when  $\angle(\mathbf{a}_1, \mathbf{a}_N) = 30^\circ$ , we can see that the optimum batch length is different for  $T = 5$  than for  $T = 1$ .

### D. Blind Speech Extraction of a Moving Speaker: A Case Study

To demonstrate the applicability of the proposed method, we consider a speech enhancement task where the speaker is moving. A six-channel recording of a speaker uttering in a multi-source noisy environment is taken from the CHiME-4 challenge database.<sup>10</sup> [55]. One-unit FastDIVA is applied in the short-term Fourier transform domain (the window length is 512 samples and the hop size is 128) in order to extract the

<sup>10</sup>The presented utterance is F04\_053C010W\_BUS.WAV

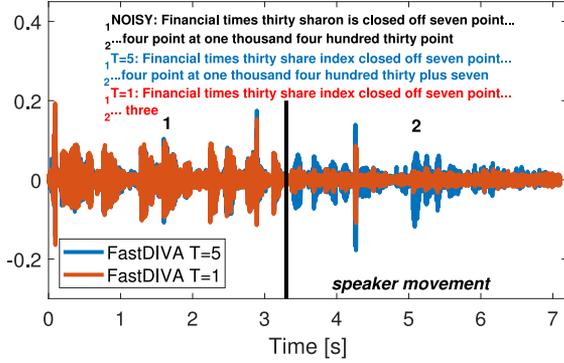


Fig. 8. Extracted utterances from a noisy recording of a moving speaker. The ground truth transcription is “Financial times thirty share index closed off seven point four points at one thousand four hundred thirty point seven”; NOISY stands for the automatic transcription from the original noisy recording by the first microphone;  $T = 1$  and  $T = 5$  correspond to the One-unit FastDIVA outputs considering, respectively, the static and the CSV mixing model.

speech. Fig. 8 shows the resulting signals and compares the ground truth transcription with automatic transcriptions by the Google Speech-to-Text system<sup>11</sup>.

Within this particular recording, the speaker moves out of its initial position in the interval 3.3 – 7.1 s. One-unit FastDIVA with  $T = 1$  (static mixing model) focuses only on the initial speaker position. This causes that the extracted voice is vanishing during the interval of the movement; the corresponding part of the automatic transcription is therefore erroneous. When the algorithm is used with  $T = 5$ , the whole utterance is successfully extracted, which results in a significantly more accurate transcription.

## VI. CONCLUSIONS

In this paper, we propose powerful BSS algorithms suitable for separating dynamic CSV-separable mixtures where the avoidance of the discontinuity problem is guaranteed. Joint separation similar to IVA, which helps us solve the permutation ambiguity, is considered as well. One-unit FastDIVA has been shown as effective for the BSE when the SOI obeys the CSV mixing model. The performance analysis has been derived for a general model pdf of the SOI, and it has been proven that One-unit FastDIVA attains the Cramér-Rao lower bound asymptotically when the background is circular Gaussian and the model pdf corresponds to the true one. Symmetric and block-deflation FastDIVA have been validated in the problem of separating several signals from CSV-separable mixtures obeying condition (C1) and (C2), respectively. The results of experiments indicate that the algorithms achieve superior interference-to-signal ratio compared to methods assuming the conventional static mixing model, especially, when the time-variability of the mixture is mild. The reliability of the separation can be supported by a proper initialization; other forms of partial knowledge about the mixing parameters might be considered in future works, as in [10].

<sup>11</sup>The transcriptions were performed by the system on December 8, 2020 at <https://cloud.google.com/speech-to-text>

By generalizing the mixing model, we have touched on the very basis of the problem that ICA and IVA solve. For this reason, many theoretical and practical questions arise. In particular, the question is what mixtures can be approximated well enough by a CSV-separable model with  $T \ll N$ . Equivalently: Which of the signals in (2) can be extracted based on the CSV model and what does this mean in practice? The experiments here and elsewhere [23], [24] show that the mixture can be dynamic only to a limited extent. For example, it is better when only some sources are moving. In addition, their movements (within the processed batch of data) should be spatially limited so that a separating vector that covers the entire motion space exists. That space should not be intersected by the motion trajectories of the other sources. A more specific analysis is the subject of further research.

Finally, it is worth pointing to the fact that the static ICA/IVA problem has the property that if at most one of the sources is Gaussian and the others are non-Gaussian, then there are no false solutions. That is, there are no independent components that do not correspond to the original signals except for the scale and order [56]. For  $T > 1$ , similar analysis does not exist yet.

## APPENDIX A

### PROOF OF PROPOSITION 1

By considering  $N = +\infty$ , all estimated values and averages are replaced by the true expectation values. Using the complex derivative identities [45], it holds that  $\mathcal{C}_{OG}$  is a real function, so  $\frac{\partial}{\partial \mathbf{w}^H} \mathcal{C}_{OG} = (\frac{\partial}{\partial \mathbf{w}^T} \mathcal{C}_{OG})^*$ , and by definition (29)

$$\mathbf{H}_1 = \frac{\partial^2 \mathcal{C}_{OG}}{\partial \mathbf{w}^T \partial \mathbf{w}} = \frac{\partial \nabla^H}{\partial \mathbf{w}} = \left[ \frac{\partial}{\partial \mathbf{w}^H} \left( \mathbf{a}^T - \nu^{-1} \mathbb{E} \left[ \phi \frac{\mathbf{x}^T}{\sigma} \right] \right) \right]^* \quad (96)$$

$$\mathbf{H}_2 = \frac{\partial^2 \mathcal{C}_{OG}}{\partial \mathbf{w}^H \partial \mathbf{w}} = \frac{\partial \nabla^T}{\partial \mathbf{w}} = \frac{\partial}{\partial \mathbf{w}} \left( \mathbf{a}^T - \nu^{-1} \mathbb{E} \left[ \phi \frac{\mathbf{x}^T}{\sigma} \right] \right), \quad (97)$$

where  $\phi(\frac{s}{\sigma})$  is, for brevity, written without the argument (which is always  $\frac{s}{\sigma}$ ). Note that the dependent variables on  $\mathbf{w}$  are  $s = \mathbf{w}^H \mathbf{x}$ ,  $\mathbf{a}$  through the OGC, and  $\sigma$  through (11);  $\nu$  is treated as a constant in (96) and (97). Using the following auxiliary expressions,

$$\frac{\partial \mathbf{a}^T}{\partial \mathbf{w}} = \frac{\mathbf{C}^*}{\sigma^2} - \mathbf{a}^* \mathbf{a}^T \quad \frac{\partial \mathbf{a}^T}{\partial \mathbf{w}^H} = -\mathbf{a} \mathbf{a}^T, \quad (98)$$

$$\frac{\partial}{\partial \mathbf{w}} \frac{1}{\sigma} = -\frac{\mathbf{a}^*}{2\sigma} \quad \frac{\partial}{\partial \mathbf{w}^H} \frac{1}{\sigma} = -\frac{\mathbf{a}}{2\sigma}, \quad (99)$$

$$\frac{\partial}{\partial \mathbf{w}} \frac{s^*}{\sigma} = \frac{\mathbf{x}^*}{\sigma} - \frac{s^* \mathbf{a}^*}{2\sigma} \quad \frac{\partial}{\partial \mathbf{w}^H} \frac{s}{\sigma} = \frac{\mathbf{x}}{\sigma} - \frac{s \mathbf{a}}{2\sigma}, \quad (100)$$

straightforward computations give

$$\begin{aligned} \frac{\partial}{\partial \mathbf{w}^H} \phi \frac{\mathbf{x}^T}{\sigma} &= \frac{\partial \phi}{\partial s} \left( \frac{\mathbf{x}}{\sigma} - \frac{s \mathbf{a}}{2\sigma} \right) \frac{\mathbf{x}^T}{\sigma} - \frac{\partial \phi}{\partial s^*} \frac{s^* \mathbf{a}^* \mathbf{x}^T}{2\sigma \sigma} - \phi \frac{\mathbf{a} \mathbf{x}^T}{2\sigma}, \\ \frac{\partial}{\partial \mathbf{w}} \phi \frac{\mathbf{x}^T}{\sigma} &= \frac{\partial \phi}{\partial s^*} \left( \frac{\mathbf{x}^*}{\sigma} - \frac{s^* \mathbf{a}^*}{2\sigma} \right) \frac{\mathbf{x}^T}{\sigma} - \frac{\partial \phi}{\partial s} \frac{s \mathbf{a}^* \mathbf{x}^T}{2\sigma \sigma} - \phi \frac{\mathbf{a}^* \mathbf{x}^T}{2\sigma}. \end{aligned}$$

By taking the expectation values of the latter expressions and using the fact that  $\mathbf{x} = \mathbf{a}\mathbf{s} + \mathbf{y}$  where  $\mathbf{s}$  and  $\mathbf{y}$  have zero mean values and are independent, we obtain

$$\frac{\partial}{\partial \mathbf{w}^H} \mathbb{E} \left[ \phi \frac{\mathbf{x}^T}{\sigma} \right] = \frac{1}{2} (\eta - \xi - \nu) \mathbf{a} \mathbf{a}^T + \frac{\rho}{\sigma^2} \mathbf{P}_y, \quad (101)$$

$$\frac{\partial}{\partial \mathbf{w}} \mathbb{E} \left[ \phi \frac{\mathbf{x}^T}{\sigma} \right] = \frac{1}{2} (\xi - \eta - \nu) \mathbf{a}^* \mathbf{a}^T + \frac{\rho}{\sigma^2} \mathbf{C}_y^*, \quad (102)$$

where  $\mathbf{P}_y = \mathbb{E}[\mathbf{y}\mathbf{y}^T]$  is the pseudo-covariance of  $\mathbf{y}$ , which is zero due to the assumption of circularity of the background signals. Putting (101) with  $\mathbf{P}_y = \mathbf{0}$  and (98) into (96), we get (30).

Finally, note that  $\mathbf{C} = \mathbf{a}\mathbf{a}^H \sigma^2 + \mathbf{C}_y$ , so (102) is equal to  $\frac{\partial}{\partial \mathbf{w}} \mathbb{E}[\phi \frac{\mathbf{x}^T}{\sigma}] = (\nu c_3 - \rho) \mathbf{a}^* \mathbf{a}^T + \frac{\rho}{\sigma^2} \mathbf{C}^*$ , where  $c_3$  is defined by (34). Putting this and (98) into (97), and using definitions (32) and (33), we get (31).

## APPENDIX B PROOF OF PROPOSITION 2

Applying the Woodbury identity to (31) gives

$$\mathbf{H}_2^{-1} = \frac{1}{c_1} \left( \mathbf{C}^{-1} - \mathbf{C}^{-1} \mathbf{a} \left( \frac{c_1}{c_2} + \mathbf{a}^H \mathbf{C} \mathbf{a} \right)^{-1} \mathbf{a}^H \mathbf{C}^{-1} \right)^T. \quad (103)$$

Using the following equalities due to the OGC imposed between  $\mathbf{a}$  and  $\mathbf{w}$ ,

$$\mathbf{C} \mathbf{w} (\mathbf{a}^H \mathbf{C}^{-1} \mathbf{a}) = \mathbf{a}, \quad \mathbf{w}^H \mathbf{a} = 1, \quad (104)$$

$$\mathbf{a}^H \mathbf{C}^{-1} \mathbf{a} = \frac{1}{\sigma^2}, \quad \mathbf{C}^{-1} \mathbf{a} = \frac{\mathbf{w}}{\sigma^2}, \quad (105)$$

(103) can be written as

$$\mathbf{H}_2^{-1} = \frac{1}{c_1} \left( \mathbf{C}^{-1} - \frac{c_2}{\sigma^2(\sigma^2 c_1 + c_2)} \mathbf{w} \mathbf{w}^H \right)^T, \quad (106)$$

and since  $\mathbf{H}_1^* = c_3 \mathbf{a} \mathbf{a}^T$ , after simplifications,

$$\mathbf{H}_1^* \mathbf{H}_2^{-1} = -\mathbf{a} \mathbf{w}^T \quad (107)$$

$$\mathbf{H}_1^* \mathbf{H}_2^{-1} \mathbf{H}_1 = -c_3^* \mathbf{a} \mathbf{a}^H \quad (108)$$

$$\mathbf{H}_2^* - \mathbf{H}_1^* \mathbf{H}_2^{-1} \mathbf{H}_1 = c_1^* (\mathbf{C} - \sigma^2 \mathbf{a} \mathbf{a}^H), \quad (109)$$

from which, by the latter equation, (39) follows.

Next, we show that  $\hat{\mathbf{H}}_1^* \hat{\mathbf{H}}_2^{-1} \nabla^* = \mathbf{0}$ . Let us denote

$$\mathbf{f} = \hat{\mathbf{E}} \left[ \phi \left( \frac{\mathbf{w}^H \mathbf{x}}{\sigma} \right) \frac{\mathbf{x}}{\sigma} \right], \quad (110)$$

which is the expression that appears in (29); so we can write that  $\nabla = \mathbf{a} - \hat{\nu}^{-1} \mathbf{f}$ . From the definition of  $\hat{\nu}$ , it follows that  $\mathbf{w}^H \mathbf{f} = \hat{\nu}$ , and by using (107),  $\hat{\mathbf{H}}_1^* \hat{\mathbf{H}}_2^{-1} \nabla^* = \mathbf{0}$ . Thus, we receive the update (40).

Let

$$\hat{\mathbf{H}}_\epsilon = \left( \frac{\hat{\nu} - \hat{\rho}}{\hat{\nu}} \right)^* \left( \frac{\hat{\mathbf{C}}}{\hat{\sigma}^2} - \epsilon \mathbf{a} \mathbf{a}^H \right), \quad (111)$$

so that  $\lim_{\epsilon \rightarrow 1} \hat{\mathbf{H}}_\epsilon = \hat{\mathbf{H}}$ , cf. (39). Using the Woodbury identity, (104), and (105), we get

$$\hat{\mathbf{H}}_\epsilon^{-1} = \left( \frac{\hat{\nu}}{\hat{\nu} - \hat{\rho}} \right)^* \left( \hat{\sigma}^2 \hat{\mathbf{C}}^{-1} + \frac{\epsilon}{\hat{\sigma}^2(1 - \epsilon)} \mathbf{w} \mathbf{w}^H \right). \quad (112)$$

Using (110),  $\nabla = \mathbf{a} - \hat{\nu}^{-1} \mathbf{f}$ ,  $\mathbf{w}^H \mathbf{f} = \hat{\nu}$ , and  $\mathbf{w}^H \nabla = 0$ . Thus

$$\hat{\mathbf{H}}_\epsilon^{-1} \nabla = \left( \frac{\hat{\nu}}{\hat{\nu} - \hat{\rho}} \right)^* \hat{\sigma}^2 \hat{\mathbf{C}}^{-1} \nabla. \quad (113)$$

The update (41) readily follows.

## REFERENCES

- [1] J. Herault and C. Jutten, "Space or time adaptive signal processing by neural network models," in *Proc. AIP Conf. Proc. 151 Neural Netw. Comput.*, (Woodbury, NY, USA), Amer. Inst. Phys. Inc., 1987, pp. 206–211.
- [2] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.
- [3] J. F. Cardoso, "Blind signal separation: Statistical principles," in *Proc. IEEE Proc. IRE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.
- [4] T.-W. Lee, *Independent Component Analysis - Theory and Applications*. Norwell, MA, USA: Kluwer Academic Publishers, 1998.
- [5] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. Hoboken, NJ, USA: Wiley, 2001.
- [6] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. Hoboken, NJ, USA: Wiley, 2002.
- [7] 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.
- [8] T. Adali, C. Jutten, A. Yeredor, A. Cichocki, and E. Moreau, "Source separation and applications [from the guest editors]," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 16–17, May 2014.
- [9] 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, p. 28, May 2018.
- [10] A. Brendel, T. Haubner, and W. Kellermann, "A unified probabilistic view on spatially informed source separation and extraction based on independent vector analysis," *IEEE Trans. Signal Process.*, vol. 68, pp. 3545–3558, 2020.
- [11] W. Liu, J. C. Principe, and S. Haykin, *Kernel Adaptive Filtering: A Comprehensive Introduction*, 1st ed. Hoboken, NJ, USA: Wiley, 2010.
- [12] M. Welling, R. S. Zemel, and G. E. Hinton, "Probabilistic sequential independent components analysis," *IEEE Trans. Neural Netw.*, vol. 15, no. 4, pp. 838–849, Jul. 2004.
- [13] 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*, May 2014, pp. 107–111.
- [14] 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.
- [15] J. Chien and H. Hsieh, "Nonstationary source separation using sequential and variational Bayesian learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 5, pp. 681–694, May 2013.
- [16] 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.
- [17] J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.
- [18] R. Mukai, H. Sawada, S. Araki, and S. Makino, *Real-Time Blind Source Separation for Moving Speech Signals*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 353–369.
- [19] F. Nesta, T. S. Wada, and B. Juang, "Batch-online semi-blind source separation applied to multi-channel acoustic echo cancellation," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 3, pp. 583–599, Mar. 2011.
- [20] M. T. Akhtar, T. Jung, S. Makeig, and G. Cauwenberghs, "Recursive independent component analysis for online blind source separation," in *Proc. IEEE Int. Symp. Circuits Syst.*, 2012, pp. 2813–2816.

- [21] A. H. Khan, M. Taseska, and E. A. P. Habets, *A Geometrically Constrained Independent Vector Analysis Algorithm for Online Source Extraction*. Liberec, Czech Republic: Springer International Publishing, 2015, pp. 396–403.
- [22] S. Hsu, T. Mullen, T. Jung, and G. Cauwenberghs, “Online recursive independent component analysis for real-time source separation of high-density eeg,” in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2014, pp. 3845–3848.
- [23] 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.*, May 2019, pp. 7903–7907.
- [24] J. Janský, Z. Koldovský, J. Málek, T. Kounovský, and J. Čmejla, “Auxiliary function-based algorithm for blind extraction of a moving speaker,” 2020, *arXiv:2002.12619*.
- [25] V. Kautský, Z. Koldovský, P. Tichavský, and V. Zarzoso, “Cramér-Rao bounds for complex-valued independent component extraction: Determined and piecewise determined mixing models,” *IEEE Trans. Signal Process.*, vol. 68, pp. 5230–5243, 2020.
- [26] 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.
- [27] 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.
- [28] T.-W. Lee, A. Bell, and R. Orglmeister, “Blind source separation of real world signals,” in *Proc. Int. Conf. Neural Netw.*, Jun. 1997, pp. 2129–2135.
- [29] T. Adali, Y. Levin-Schwartz, and V. D. Calhoun, “Multimodal data fusion using source separation: Two effective models based on ICA and IVA and their properties,” *Proc. IEEE*, vol. 103, no. 9, pp. 1478–1493, Sep. 2015.
- [30] 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.
- [31] X. Chen, Z. J. Wang, and M. McKeown, “Joint blind source separation for neurophysiological data analysis: Multiset and multimodal methods,” *IEEE Signal Process. Mag.*, vol. 33, no. 3, pp. 86–107, May 2016.
- [32] 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.
- [33] H. Sawada, R. Mukai, S. Araki, and S. Makino, “A robust and precise method for solving the permutation problem of frequency-domain blind source separation,” *IEEE Speech Audio Process.*, vol. 12, no. 5, pp. 530–538, Sep. 2004.
- [34] T. Kim, I. Lee, and T. Lee, “Independent vector analysis: Definition and algorithms,” in *Proc. 14th Asilomar Conf. Signals, Syst. Comput.*, Oct. 2006, pp. 1393–1396.
- [35] 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.
- [36] S. Gannot, D. Burshtein, and E. Weinstein, “Signal enhancement using beamforming and nonstationarity with applications to speech,” *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
- [37] V. Kautský, Z. Koldovský, and P. Tichavský, “Cramér-rao-induced bound for interference-to-signal ratio achievable through non-Gaussian independent component extraction,” in *Proc. IEEE Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, Dec. 2017, pp. 94–97.
- [38] D.-T. A. Pham, “Contrast functions for blind separation and deconvolution of sources,” in *Proc. Int. Conf. Independent Compon. Anal. Signal Separation*, Dec 2001.
- [39] D.-T. A. Pham, “Blind partial separation of instantaneous mixtures of sources,” in *Proc. Int. Conf. Independent Compon. Anal. Signal Separation*, 2006, pp. 868–875.
- [40] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2002.
- [41] Z. Koldovský, P. Tichavský, and V. Kautský, “Orthogonally constrained independent component extraction: Blind MPDR beamforming,” in *Proc. Eur. Signal Process. Conf.*, Sep. 2017, pp. 1195–1199.
- [42] H. Li and T. Adali, “Complex-valued adaptive signal processing using nonlinear functions,” *EURASIP J. Adv. Signal Process.*, vol. 2008, Feb. 2008, Art. no. 765615.
- [43] P. J. Huber, “Projection pursuit,” *Ann. Statist.*, vol. 13, pp. 435–475, Jun. 1985.
- [44] E. Bingham and A. Hyvärinen, “A fast fixed-point algorithm for independent component analysis of complex valued signals,” *Int. J. Neural Syst.*, vol. 10, pp. 1–8, Feb. 2000.
- [45] K. Kreutz-Delgado, “The complex gradient operator and the CR-calculus,” 2009, *arXiv:0906.4835*.
- [46] N. Delfosse and P. Loubaton, “Adaptive blind separation of independent sources: A deflation approach,” *Signal Process.*, vol. 45, no. 1, pp. 59 – 83, 1995.
- [47] J.-F. Cardoso, “On the performance of orthogonal source separation algorithms,” in *Proc. Eur. Signal Process. Conf.*, Sep. 1994, pp. 776–779.
- [48] B. Porat, *Digital Processing of Random Signals: Theory and Methods*. New York, NY, USA: Dover Publications, 2008.
- [49] A. Hyvärinen, “One-unit contrast functions for independent component analysis: A statistical analysis,” in *Proc. Neural Netw. Signal Process. VII. Proc. 1997 IEEE Signal Process. Soc. Workshop*, Sep. 1997, pp. 388–397.
- [50] P. Tichavský, Z. Koldovský, and E. Oja, “Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis,” *IEEE Trans. Signal Process.*, vol. 54, no. 4, pp. 1189–1203, Apr. 2006.
- [51] Z. Koldovský, V. Kautský, T. Kounovský, and J. Čmejla, “Algorithm for independent vector extraction based on semi-time-variant mixing model,” 2020, *arXiv:1910.10242*.
- [52] P. Tichavský, Z. Koldovský, and E. Oja, *Speed and Accuracy Enhancement of Linear ICA Techniques Using Rational Nonlinear Functions*, pp. 285–292. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.
- [53] B. Loesch and B. Yang, “Cramér-Rao bound for circular and noncircular complex independent component analysis,” in *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 365–379, Jan. 2013.
- [54] V. Kautský, Z. Koldovský, and P. Tichavský, “Performance bound for blind extraction of non-gaussian complex-valued vector component from Gaussian background,” in *Proc. IEEE Int. Conf. Audio, Speech Signal Process.*, May 2019, pp. 5287–5291.
- [55] E. Vincent, S. Watanabe, A. A. Nugraha, J. Barker, and R. Marxer, “An analysis of environment, microphone and data simulation mismatches in robust speech recognition,” *Comput. Speech Lang.*, vol. 46, pp. 535–557, 2017.
- [56] F. J. Theis, “Uniqueness of real and complex linear independent component analysis revisited,” in *Proc. 12th Eur. Signal Process. Conf.*, 2004, pp. 1705–1708.



**Zbyněk Koldovský** (Senior Member, IEEE) was born in Jablonec nad Nisou, Czechia, 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 in Prague, Prague, Czechia, 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, Czechia, 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, Czechia, and 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.



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

achievable performance.



**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, Czechia, 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 or 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. From 2008 to 2011 and since 2016, he has been a Member of the IEEE SPS Committee Signal Processing Theory and Methods. He was also the General Co-Chair of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing ICASSP 2011 in Prague. From 2002 and 2004, he was an Associate Editor for the IEEE SIGNAL PROCESSING LETTERS, and from 2005 to 2009 and from 2011 to 2016, for the IEEE TRANSACTIONS ON SIGNAL PROCESSING.



**Jiří Málek** (Member, IEEE) was born in Czechia, in 1983. He received the Ph.D. degree in technical cybernetics from the Technical University of Liberec, Liberec, Czechia, in 2011. Since 2011, he has been an Assistant Professor with the Faculty of Mechatronics, Technical University of Liberec. His main research interests include audio signal enhancement, blind source separation, and robust automatic speech recognition. He is a Reviewer for journals and conferences focused on digital signal processing, which include the IEEE TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING and *IET Signal Processing* or ICASSP.



**Jaroslav Čmejla** received the master's degree in 2016 in information technologies from the Faculty of Mechatronics, Informatics and Interdisciplinary Studies, Technical University of Liberec, Liberec, Czechia, where he is currently working toward the Ph.D. degree.

His main research interests include audio signal processing and blind source separation. His recent works are related to the blind source extraction problem. He is a Member of the Acoustic Signal Analysis and Processing Group led by Prof. Zbynek

Koldovsky.