Orthogonally Constrained Independent Component Extraction: Blind MPDR Beamforming

Zbyněk Koldovský*, Petr Tichavský†, and Václav Kautský‡

*Faculty of Mechatronics, Informatics, and Interdisciplinary Studies, Technical University of Liberec,
Studentská 2, 461 17 Liberec, Czech Republic. E-mail: zbynek.koldovsky@tul.cz
†Institute of Information Theory and Automation, Pod vodárenskou věží 4,
P. O. Box 18, 182 08 Praha 8, Czech Republic. E-mail: tichavsk@utia.cas.cz
‡Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague,
E-mail: kautsvac@fjfi.cvut.cz

Abstract—We propose a novel technique for the extraction of one independent component from an instantaneous linear complex-valued mixture of signals. The mixing model is optimized in terms of the number of parameters that are necessary to simultaneously estimate one column of the mixing matrix and one row of the de-mixing matrix, which both correspond to the desired source. The desired source is assumed to have a non-Gaussian distribution, while the other sources are modeled, for simplicity, as Gaussian-distributed, although in applications the other sources can be arbitrary. We propose an algorithm that can be interpreted as a blind self-steering Minimum-Power Distortionless Response (MPDR) beamformer. The method is compared with the popular Natural Gradient algorithm for general Independent Component Analysis. Their performances are comparable but the proposed method has a lower computational complexity; in examples, it is about four times faster.

I. INTRODUCTION

Consider the following complex-valued instantaneous mixture model that describes a multi-channel mixing problem with one target signal within a frequency bin [1]. The model reads

$$\mathbf{x}(t) = \mathbf{a}s(t) + \mathbf{y}(t) \qquad \text{or} \qquad \mathbf{X} = \mathbf{a}\,\mathbf{s} + \mathbf{Y} \qquad (1)$$

where the former is a vector-symbolic description while the latter describes a batch of data; t is a time or frame index which will be omitted for brevity; $\mathbf{x} = [x_1, \ldots, x_d]^T$ represents d observed signals on sensors; s represents a target signal and $\mathbf{a}$ is the corresponding $d \times 1$ *steering vector*; $\mathbf{y} = [y_1, \ldots, y_d]^T$ represents noise signals that are *independent* of s. The upper-case bold letters such as $\mathbf{X}$ and $\mathbf{Y}$ will denote matrices whose columns contain concrete samples of $\mathbf{x}$ and $\mathbf{y}$, respectively; let the number of samples be N where $N \gg d$; $\mathbf{s}$ is the $1 \times N$ row vector of samples of s. All signals are assumed to have zero mean.

In array processing literature [2], an optimum extractor of s is the Minimum Power Distortionless Response beamformer (MPDR) defined trough

$$\min_{\mathbf{w}} \mathrm{E}[|\mathbf{w}^H \mathbf{x}|^2] \qquad \text{w.r.t.} \qquad \mathbf{w}^H \mathbf{a} = 1, \qquad (2)$$

where the output signal is $\mathbf{w}^H \mathbf{x}$. The solution of (2) is

$$\mathbf{w}_{\mathrm{MPDR}} = \frac{\mathbf{C}_{\mathbf{x}}^{-1} \mathbf{a}}{\mathbf{a}^H \mathbf{C}_{\mathbf{x}}^{-1} \mathbf{a}}, \qquad (3)$$

where $\mathbf{C}_{\mathbf{x}} = \mathrm{E}[\mathbf{x}\mathbf{x}^H]$ is the covariance of $\mathbf{x}$; $\mathrm{E}[\cdot]$ stands for the expectation operator. The expression can be used in practice by replacing $\mathbf{C}_{\mathbf{x}}$ by its sample-based estimate $\widehat{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^H/N$. However, the steering vector $\mathbf{a}$ must be known.

Various models have been proposed to identify $\mathbf{a}$ from $\mathbf{x}$, see, e.g. [3]. Most of recent attempts come from Blind Source Separation (BSS) based on Independent Component Analysis (ICA) [4], [5]. In ICA, it is assumed that (1) can be written as a determined mixture

$$\mathbf{x} = \mathbf{H}\mathbf{u}, \qquad (4)$$

where $\mathbf{H} = [\mathbf{a}\,\mathbf{H}_2]$ is $d \times d$ regular, $\mathbf{H}_2$ is a $d \times (d-1)$ matrix, $\mathbf{u} = [s; \mathbf{u}_2]$ where $\mathbf{u}_2 = [u_2, \ldots, u_d]^T$ is a vector of $d-1$ independent signals. In other words, $\mathbf{y}$ is assumed to be decomposable into $d-1$ independent signals as $\mathbf{y} = \mathbf{H}_2\mathbf{u}_2$. A generalization to ICA is Multidimensional ICA (MICA) [6], which enables the components of $\mathbf{u}_2$ to be dependent. Blind Source Extraction (BSE) is a class of methods that aims to extract one signal of specific properties (not necessarily an independent signal); see, e.g., [7]. ICA, MICA and BSE are bases of further extensions as is Independent Vector Analysis (IVA) [8] or Joint BSS (JBSS) [9]; see also [10] for further extensions and perspectives.

Since the role of s can be interchanged with any independent component of $\mathbf{y}$, a priori knowledge is needed in order to separate the desired source. Typically, information is given in the form of a rough estimate of $\mathbf{a}$, e.g., based on an approximate direction-of-arrival (DOA) of the target source on the sensors [11]. An ICA (MICA) algorithm can be then initialized so that it converges to the desired solution.

In this paper, we focus on this situation when an initial guess for the steering vector is given, and the nearest independent source should be found. ICA as well as MICA appear to be too excessive for this goal, because they also aim to separate $\mathbf{y}$. Therefore, in Section II, we propose a novel mixing model

where the separation of $\mathbf{y}$ is not the goal. We will refer to it as to *Independent Component Extraction (ICE)*[1].

The rest of the paper is organized as follows. In Section III, an initialization-dependent algorithm for the extraction of a nonGaussian target source is derived based on maximum likelihood principle. While s is assumed to be nonGaussian, $\mathbf{y}$ is modeled as a mixture of correlated Gaussian signals. The algorithm is faster than the popular Natural Gradient algorithm by Amari et al. [13], which is shown by simulations in Section IV. Section V concludes the paper.

II. MIXING MODEL PARAMETRIZATION

A. Mixing and De-mixing Matrices

The model (1) involves the scaling ambiguity that the couple $(s, \mathbf{a})$ can be replaced by $(\alpha s, \alpha^{-1}\mathbf{a})$ for any $\alpha \neq 0$. Without any loss of generality, we can cope with this problem by putting a_1, the first element of $\mathbf{a}$, equal to one; let us denote[2]

$$\mathbf{a} = [1; \mathbf{g}]. \tag{5}$$

This is equivalent to considering s as the signal of the target source as it is received by the first sensor. In multi-microphone systems, the elements of $\mathbf{g}$ are, after this normalization, referred to as Relative Transfer Functions (RTFs); see [14].

The ICE model describes (1) as a determined mixture, so the subspace spanned by $[s; \mathbf{y}]$ must have dimension d. In other words, $\mathbf{y}$ is assumed to be coherent so that it can be fully separated from $\mathbf{x}$. Hence, $\mathbf{y}$ can be expressed as $\mathbf{y} = \mathbf{Qz}$ where $\mathbf{Q}$ is a $d \times (d-1)$ matrix and $\mathbf{z} = [z_1, \ldots, z_{d-1}]$.

Let the mixing and the de-mixing descriptions of (1) with square matrices be, respectively,

$$\mathbf{x} = \mathbf{Av} \qquad \text{and} \qquad \mathbf{v} = \mathbf{Wx}, \tag{6}$$

where $\mathbf{v} = [s; \mathbf{z}]$. The structure of the mixing matrix is therefore $\mathbf{A} = [\mathbf{a}, \mathbf{Q}]$. Let the de-mixing matrix be partitioned as $\mathbf{W} = [\mathbf{w}^H; \mathbf{B}]$. The role of $\mathbf{w}$ is the separating vector that extracts s from $\mathbf{x}$. By contrast, $\mathbf{B}$ extracts $\mathbf{z}$ while it blocks out s, so its role is the same as that of "Blocking Matrix" in array processors [2], [14]. This property is ensured whenever $\mathbf{B}$ is orthogonal to $\mathbf{a}$, so we put $\mathbf{B} = [\mathbf{g}, -\mathbf{I}_{d-1}]$, by which the definition of $\mathbf{z}$ is completed, i.e., $\mathbf{z} = \mathbf{By} = \mathbf{Bx}$; we will refer to $\mathbf{z}$ as to the *background* signals.

Since $\mathbf{w}^H\mathbf{x} = s$, the vector $\mathbf{w}$ represents a distortionless beamformer satisfying constraint

$$\mathbf{w}^H\mathbf{a} = 1. \tag{7}$$

By parameterizing the vector as $\mathbf{w}^H = [\beta, \mathbf{h}^H]$, it follows from (5) and (7) that $\beta = 1 - \mathbf{h}^H\mathbf{g}$.

Now, by stating that $\mathbf{WA} = \mathbf{I}_d$, the concrete parametrization of the mixing and the de-mixing matrix, respectively, is

$$\mathbf{A} = [\mathbf{a}, \mathbf{Q}] = \begin{pmatrix} 1 & \mathbf{h}^H \\ \mathbf{g} & \mathbf{g}\mathbf{h}^H - \mathbf{I}_{d-1} \end{pmatrix}, \tag{8}$$

[1]At the time of submission of this paper, a similar idea appeared in [12], however, for situations where the steering vector $\mathbf{a}$ is known in advance. Otherwise, the idea is novel to the best of our knowledge.

[2]Throughout the paper, we use Matlab convention for matrix/vector concatenation, e.g., $[1; \mathbf{g}] = [1, \mathbf{g}^T]^T$.

and

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}^H \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} 1 - \mathbf{h}^H\mathbf{g} & \mathbf{h}^H \\ \mathbf{g} & -\mathbf{I}_{d-1} \end{pmatrix}. \tag{9}$$

It should be emphasized that the ICE model is minimized for the problem of separating s from $\mathbf{x}$. Although we do not care about the form in which $\mathbf{y}$ is separated from $\mathbf{x}$, the extracted counterpart of $\mathbf{y}$, the background signal $\mathbf{z}$, is fully specified. The mixing model has only $2d - 2$ parameters represented by the elements of $\mathbf{g}$ and $\mathbf{h}$.

This is in contrast with MICA, which is highly ambiguous in that $\mathbf{Q}$ and $\mathbf{z}$ can be arbitrary but $\mathbf{y} = \mathbf{Qz}$. In ICA, the elements of $\mathbf{z}$ should be independent, which is less ambiguous as compared to MICA but still ambiguous and limiting as compared to ICE.

B. Orthogonal Constraint

Some ICA and MICA algorithms impose orthogonal constraint (OG) on the separated signals [15]. The separated signal subspaces are constrained to be orthogonal to ensure their uncorrelatedness. The constraint reduces the number of ambiguous parameters in the mixing model.

Let $\mathbf{W}$ denote a de-mixing matrix estimate and

$$\widehat{\mathbf{V}} = \begin{pmatrix} \widehat{\mathbf{s}} \\ \widehat{\mathbf{Z}} \end{pmatrix} = \mathbf{WX} \tag{10}$$

be concrete samples of the estimated de-mixed signals. The OG requires that

$$\widehat{\mathbf{s}} \cdot \widehat{\mathbf{Z}}^H / N = \mathbf{w}^H \widehat{\mathbf{C}}_{\mathbf{x}} \mathbf{B}^H = \mathbf{0}, \tag{11}$$

where $\widehat{\mathbf{C}}_{\mathbf{x}} = \mathbf{XX}^H / N$ is the sample-based estimate of $\mathbf{C}_{\mathbf{x}}$. To simplify the notation, we do not explicitly write the dependence of $\mathbf{B}$, $\mathbf{w}$, $\mathbf{a}$, etc. on $\mathbf{g}$ and $\mathbf{h}$.

Once OG is applied, there is a coupling between $\mathbf{g}$ and $\mathbf{h}$, which reduces the number parameters from $2d - 2$ to $d - 1$. Let the free variable be $\mathbf{g}$. It can be shown that $\mathbf{g}$ and $\mathbf{h}$ are coupled through

$$\mathbf{w}_{\text{OG}} = \frac{\widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\mathbf{a}}{\mathbf{a}^H\widehat{\mathbf{C}}_{\mathbf{x}}^{-1}\mathbf{a}}. \tag{12}$$

By comparing (12) with (3), $\mathbf{w}_{\text{OG}}$ can be interpreted as an approximate MPDR beamformer where $\mathbf{C}_{\mathbf{x}}$ is replaced by $\widehat{\mathbf{C}}_{\mathbf{x}}$, and $\mathbf{a}$ is replaced by its current estimate given by $[1; \mathbf{g}]$.

III. PROPOSED ALGORITHM

In this paper, we will focus on ICE constrained by the OG. The task can be formulated as follows.

Definition 1. *The goal of the orthogonally constrained ICE is to find* $\mathbf{a}$, *or, equivalently* $\mathbf{g}$, *such that* $\widehat{\mathbf{s}} = \mathbf{w}_{\text{OG}}^H\mathbf{X}$ *is as independent of its orthogonal complement* $\widehat{\mathbf{Z}} = \mathbf{BX}$ *as possible.*

There are several statistical models in ICA relying on one or more signals' properties such as nonGaussianity, nonstationarity or nonwhiteness [16]. In this paper, we focus on the nonGaussianity-based model where all signals are modeled as i.i.d. sequences. We will constrain to a particular case where

1) the target signal s is drawn from a nonGaussian distribution, while

2) the background signals $\mathbf{z}$ are circularly-symmetric Gaussian.

The latter assumption is not that restrictive as it might seem. In the ICE de-mixing model, $\mathbf{z} = \mathbf{Bx}$ correspond to "mixed" and correlated signals (up to very special cases). Due to the Central Limit Theorem, mixed signals tend to be more Gaussian [17]. Moreover, the multivariate Gaussian distribution is the simplest statistical model of correlated signals.

A. Maximum Likelihood Principle

Let $p_s(\xi_1)$ and $p_\mathbf{z}(\boldsymbol{\xi}_2)$ denote the probability density functions of s and $\mathbf{z}$, respectively; ξ_1 and $\boldsymbol{\xi}_2$ denote free variables of appropriate dimensions. Using the independence assumption, the joint pdf of s and $\mathbf{z}$ is

$$p_\mathbf{s}(\boldsymbol{\xi}) = p_s(\xi_1) \cdot p_\mathbf{z}(\boldsymbol{\xi}_2), \tag{13}$$

where $\boldsymbol{\xi} = [\xi_1; \boldsymbol{\xi}_2]$, and $p_\mathbf{z}(\boldsymbol{\xi}_2) \sim \mathcal{CN}(\mathbf{0}, \mathbf{C_z})$. From (6) and (9), the joint pdf of $\mathbf{x}$ is

$$p_\mathbf{x}(\boldsymbol{\xi}) = p_s(\mathbf{w}^H \boldsymbol{\xi}) \cdot p_\mathbf{z}(\mathbf{B}\boldsymbol{\xi}) \cdot \underbrace{\det \mathbf{W}}_{=1} \tag{14}$$

$$= p_s((1 - \mathbf{h}^H \mathbf{g})\xi_1 + \mathbf{h}^H \boldsymbol{\xi}_2) \cdot p_\mathbf{z}(\xi_1 \mathbf{g} - \boldsymbol{\xi}_2). \tag{15}$$

For the measured signals $\mathbf{X}$, the log-likelihood function, as a function of the model parameters $\mathbf{g}$ and $\mathbf{h}$, reads

$$\mathcal{L}(\mathbf{g}, \mathbf{h}) = \frac{1}{N} \log \prod_{i=1}^{N} p_\mathbf{x}(\mathbf{g}, \mathbf{h}|\mathbf{x}_i) \tag{16}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(\log p_s(\mathbf{w}^H \mathbf{x}_i) + \log p_\mathbf{z}(\mathbf{Bx}_i) \right) \tag{17}$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left(\log p_s(\mathbf{w}^H \mathbf{x}_i) - \mathbf{x}_i^H \mathbf{B}^H \mathbf{C_z}^{-1} \mathbf{Bx}_i \right) + C, \tag{18}$$

where $\mathbf{x}_i$ denotes the ith column of $\mathbf{X}$, that is, the ith sample of $\mathbf{x}$, and C is a constant independent of $\mathbf{g}$ and $\mathbf{h}$. By applying the OG and replacing $\mathbf{C_z}$ by $\widehat{\mathbf{C}}_\mathbf{z} = \widehat{\mathbf{Z}}\widehat{\mathbf{Z}}^H/N$, we introduce a contrast function for $\mathbf{g}$ as

$$\mathcal{C}(\mathbf{g}) = \frac{1}{N} \sum_{i=1}^{N} \left(\log p_s(\mathbf{w}_{\mathrm{OG}}^H \mathbf{x}_i) - \mathbf{x}_i^H \mathbf{B}^H \widehat{\mathbf{C}}_\mathbf{z}^{-1} \mathbf{Bx}_i \right). \tag{19}$$

The gradient of (19) with respect to $\mathbf{g}^H$ is

$$\frac{\partial \mathcal{C}(\mathbf{g})}{\partial \mathbf{g}^H} = \frac{1}{N} \sum_{i=1}^{N} \Big(\lambda_\mathbf{a} \mathcal{E} \Big\{ \widehat{\mathbf{C}}_\mathbf{x}^{-1} \big(2\Re\{\overline{\psi_s(\widehat{s}_i)}\}(\widehat{s}_i)\big) \mathbf{a}$$
$$- \overline{\psi_s(\widehat{s}_i)}\mathbf{x}_i \Big\} - \overline{x}_{1i}\widehat{\mathbf{C}}_\mathbf{z}^{-1}\mathbf{Bx}_i \Big), \tag{20}$$

where $\widehat{s}_i = \mathbf{w}_{\mathrm{OG}}^H \mathbf{x}_i$, $\lambda_\mathbf{a} = (\mathbf{a}^H \widehat{\mathbf{C}}_\mathbf{x}^{-1}\mathbf{a})^{-1}$, $\mathcal{E}\{\cdot\}$ eliminates the first element of vector-valued argument, $\Re\{\cdot\}$ is the real part of the argument, x_{1i} is the first element of $\mathbf{x}_i$, and

$$\psi_s(\xi_1) = -\frac{\partial \log p_s(\xi_1)}{\partial \overline{\xi}_1} \tag{21}$$

Algorithm 1: Gradient ascent algorithm for OG constrained Gaussian background ICE: OGICE

Input: $\mathbf{X}$, $\mathbf{g}_{\mathrm{ini}}$, μ, `tol`
Output: $\mathbf{a}, \mathbf{w}$

1   $\widehat{\mathbf{C}}_\mathbf{x} = \mathbf{XX}^H/N$;
2   $\mathbf{g} = \mathbf{g}_{\mathrm{ini}}$;
3 **repeat**
4     $\mathbf{a} \leftarrow [1; \mathbf{g}]$;   /* according to (5) */
5     $\lambda_\mathbf{a} \leftarrow (\mathbf{a}^H \widehat{\mathbf{C}}_\mathbf{x}^{-1}\mathbf{a})^{-1}$;
6     $\mathbf{w} \leftarrow \lambda_\mathbf{a}\widehat{\mathbf{C}}_\mathbf{x}^{-1}\mathbf{a}$;   /* OG constraint (12) */
7     $\widehat{\mathbf{s}} \leftarrow \mathbf{w}^H\mathbf{X}$;
8     $\mathbf{q} \leftarrow \phi(\widehat{\mathbf{s}})/(\widehat{\mathbf{s}}\phi(\widehat{\mathbf{s}})^H/N)$;   /* to fulfil (26) */
9     $\Delta \leftarrow \mathbf{h} - \mathcal{E}\{\lambda_\mathbf{a}\widehat{\mathbf{C}}_\mathbf{x}^{-1}\mathbf{Xq}^H/N\}$;   /* by (27) */
10    $\mathbf{g} \leftarrow \mathbf{g} + \mu\Delta$;   /* gradient ascent */
11 **until** $\|\Delta\| < $ `tol`;

---

is the score function corresponding to $p_s(\cdot)$. The last term in (20) can be simplified according to

$$\frac{1}{N} \sum_{i=1}^{N} \overline{x}_{1i}\widehat{\mathbf{C}}_\mathbf{z}^{-1}\mathbf{Bx}_i = \widehat{\mathbf{C}}_\mathbf{z}^{-1}\widehat{\mathbf{Z}}\mathbf{X}^H\mathbf{e}_1/N = \tag{22}$$

$$= \widehat{\mathbf{C}}_\mathbf{z}^{-1}\widehat{\mathbf{Z}}\big[\widehat{\mathbf{s}}^H, \quad \widehat{\mathbf{Z}}^H\big]\mathbf{A}^H\mathbf{e}_1/N = \tag{23}$$

$$= \big[\underbrace{\widehat{\mathbf{C}}_\mathbf{z}^{-1}\widehat{\mathbf{Z}}\widehat{\mathbf{s}}}_{=\mathbf{0}}^H/N, \quad \mathbf{I}_{d-1}\big]\mathbf{A}^H\mathbf{e}_1 = \mathbf{h}, \tag{24}$$

where $\mathbf{e}_1$ denotes the first column of $\mathbf{I}_d$, and $\widehat{\mathbf{s}} = \mathbf{w}_{\mathrm{OG}}^H\mathbf{X}$. The fact that $\widehat{\mathbf{Z}}\widehat{\mathbf{s}}^H = \mathbf{0}$ follows from the OG (11), and the last equation follows from (8).

B. Gradient Ascent Algorithm

In practice, the pdf of the target signal is not known, so, in ICA, the score function (21) is replaced by an appropriate nonlinearity $\phi(\cdot)$. For ICE, an additional requirement is that ϕ satisfies that

$$\mathrm{E}[s\overline{\phi(s)}] = 1. \tag{25}$$

It could be easily verified that for $\phi = \psi_s$, this condition is satisfied. By inspecting (20) for $N \to +\infty$, it is seen that the gradient is zero when $\mathbf{a}$ is the true steering vector, so the optimum solution is the extreme of the contrast function. To preserve this key property even when $\phi \neq \psi_s$, the condition (25) must be fulfilled.

Since (25) cannot be ensured without knowing p_s, we propose to normalize the selected nonlinearity ϕ so that the sample-based moment

$$\widehat{\mathbf{s}}\phi(\widehat{\mathbf{s}})^H/N = 1, \tag{26}$$

where the scalar function ϕ is applied element-wise.

Now, using (24) and (26) and the fact that $\mathcal{E}\{\lambda_\mathbf{a}\widehat{\mathbf{C}}_\mathbf{x}^{-1}\mathbf{a}\} = \mathbf{h}$, the gradient (20) simplifies to

$$\frac{\partial \mathcal{C}}{\partial \mathbf{g}^H} = \mathbf{h} - \mathcal{E}\{\lambda_\mathbf{a}\widehat{\mathbf{C}}_\mathbf{x}^{-1}\mathbf{X}\phi(\widehat{\mathbf{s}})^H/N\}. \tag{27}$$

Based on this, we propose the gradient ascent algorithm described in Algorithm 1, which will be referred to as OGICE.

The method starts from an initial guess $\mathbf{g}_{\text{ini}}$ and performs small steps in the direction of the gradient (27). The length of the step is controlled through a positive constant μ. In each step, the nonlinearity ϕ is normalized to satisfy (26). The algorithm stops when the norm of the gradient is smaller than a selected tolerance.

C. OGICE vs. (Natural) Gradient-Based ICA

Consider a special case where $\widehat{\mathbf{C}}_{\mathbf{x}} = \mathbf{I}_d$ (which can be achieved by de-correlating the input data $\mathbf{X}$ as in Algorithm 2). Then, $\mathbf{w}_{\text{OG}} = \mathbf{a}/\|\mathbf{a}\|^2$ and (27) can be written in the form

$$\|\mathbf{a}\|^2 \frac{\partial \mathcal{C}}{\partial \mathbf{g}^H} = \mathbf{g} - \mathcal{E}\{\mathbf{X}\phi(\widehat{\mathbf{s}})^H/N\}. \tag{28}$$

This reveals an interesting similarity between OGICE and the method by Bell and Sejnowski [18] for the classical ICA formulation, whose popular improvement is known as the Natural Gradient algorithm [13], [19]. The update rule of [18] for the (ICA) de-mixing matrix is

$$\Delta \mathbf{W} \leftarrow \mathbf{W}^{-H} - \phi(\mathbf{W}^H \mathbf{X})\mathbf{X}^H/N. \tag{29}$$

After the conjugate transpose and by denoting $\mathbf{W}^{-1} = \mathbf{A}$, the update rule is

$$\Delta \mathbf{W}^H \leftarrow \mathbf{A} - \mathbf{X}\phi(\mathbf{W}^H \mathbf{X})^H/N. \tag{30}$$

Now, the right-hand side of (28) corresponds to any column of the right-hand side of (30). Nevertheless, there are principal differences between the related algorithms. In (30), the update of any row of $\mathbf{W}$ is influenced by the other rows (the estimation of one component influences that of the others), because $\mathbf{A} = \mathbf{W}^{-1}$. Next, (28) is used under the condition (26) and when the data are de-correlated, which is not required by the ICA algorithm using (30).

Algorithm 2: OGICE with preprocessing

Input: $\mathbf{X}, \mathbf{g}_{\text{ini}}, \mu, \texttt{tol}$
Output: $\mathbf{a}, \mathbf{w}$
1   $\widehat{\mathbf{C}}_{\mathbf{x}} = \mathbf{X}\mathbf{X}^H/N; \mathbf{D} = \widehat{\mathbf{C}}_{\mathbf{x}}^{-1/2};$
2   $\mathbf{X} \leftarrow \mathbf{D}\mathbf{X};$ $/\ast$ de-correlation $\ast/$
3   $\mathbf{a}_{\text{ini}} \leftarrow \mathbf{D}[1; \mathbf{g}_{\text{ini}}]; \alpha = (\mathbf{a}_{\text{ini}})_1;$
4   $\mathbf{g} \leftarrow \mathcal{E}\{\mathbf{a}_{\text{ini}}/\alpha\};$
5   **repeat**
6      $\mathbf{a} \leftarrow [1; \mathbf{g}];$
7      $\mathbf{w} \leftarrow \mathbf{a}/\|\mathbf{a}\|^2;$
8      $\widehat{\mathbf{s}} \leftarrow \mathbf{w}^H \mathbf{X};$
9      $\mathbf{q} \leftarrow \phi(\widehat{\mathbf{s}})/(\widehat{\mathbf{s}}\phi(\widehat{\mathbf{s}})^H/N);$
10     $\Delta \leftarrow \mathcal{E}\{\mathbf{a} - \mathbf{X}\mathbf{q}^H/N\}/\|\mathbf{a}\|^2;$
11     $\mathbf{g} \leftarrow \mathbf{g} + \mu\Delta;$
12   **until** $\|\Delta\| < \texttt{tol};$
13   $\mathbf{a} \leftarrow \alpha\mathbf{D}^{-1}\mathbf{a}; \mathbf{w} \leftarrow \overline{\alpha}^{-1}\mathbf{D}\mathbf{w};$
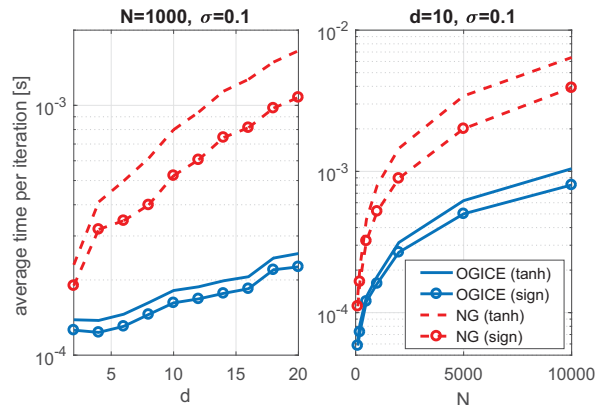

Fig. 1. Average computational time per iteration for varying d and N.

IV. EXPERIMENTAL RESULTS

In simulations, we compared OGICE with the popular Natural Gradient (NG) algorithm from [13]. Both methods were initialized using the true steering vector perturbed by random errors taken from $\mathcal{CN}(0, \sigma^2)$. In NG, the initial de-mixing matrix consists of the first row equal to (12) while the other rows are selected as in (9). In both methods, two nonlinearities $\phi(\cdot)$ were tested. The first is $\tanh(\cdot)$, which was employed by many ICA methods [4]. The second nonlinearity is $\text{sign}(x) = x/|x|$, which corresponds to the score function of the circularly-symmetric complex Laplacian distribution. To compare, we consider also the performance of the approximate MPDR beamformer (using sample-based covariance matrix) (12) where $\mathbf{a}$ is either equal to the true value (oracle) or to its perturbed value (initial).

The target signal was generated from the latter distribution with variance one. The (interference) signals were generated as circularly-symmetric Gaussian $\mathcal{CN}(\mathbf{0}, \mathbf{I}_{d-1})$. The elements of the mixing matrix $\mathbf{A}$ were generated independently from $\mathcal{CN}(1, 1/4)$ (the mean value is one to simulate similarly distributed signals over the sensors); we set $(\mathbf{A})_{11} = 1$.

The signals were mixed according to $\mathbf{X} = \mathbf{A}\mathbf{S}$. This corresponds to (1) where $\mathbf{a}$ is equal to the first row of $\mathbf{A}$ while the covariance of the noise signals $\mathbf{y}$ is $\mathbf{C}_{\mathbf{y}} = \mathbf{A}_2\mathbf{A}_2^H$, where $\mathbf{A}_2$ consists of columns $2, \dots, d$ of $\mathbf{A}$. In each trial, the signals were generated, mixed and separated by the algorithms. The separated signal was evaluated in terms of Signal-to-Interference Ratio (SIR). Fig. 2 shows the SIR averaged over 10^3 trials, respectively, for varying d, N, and σ when the other parameters were fixed to $d = 10$, $N = 10^3$ and $\sigma = 0.1$; $\mu = 0.01$ was used in OGICE as well as in NG.

The influence of the nonlinearity $\phi(\cdot)$ on SIR is significant. OGICE as well as NG yields better SIR when $\phi(\cdot)$ corresponds to the true score function $\text{sign}(\cdot)$ of the target signal distribution. NG appears to be more sensitive to the choice of $\phi(\cdot)$. While it achieves almost superior results with $\text{sign}(\cdot)$, it performs worst with $\tanh(\cdot)$. The fact that the approximate MPDR with oracle steering vector can be outperformed by
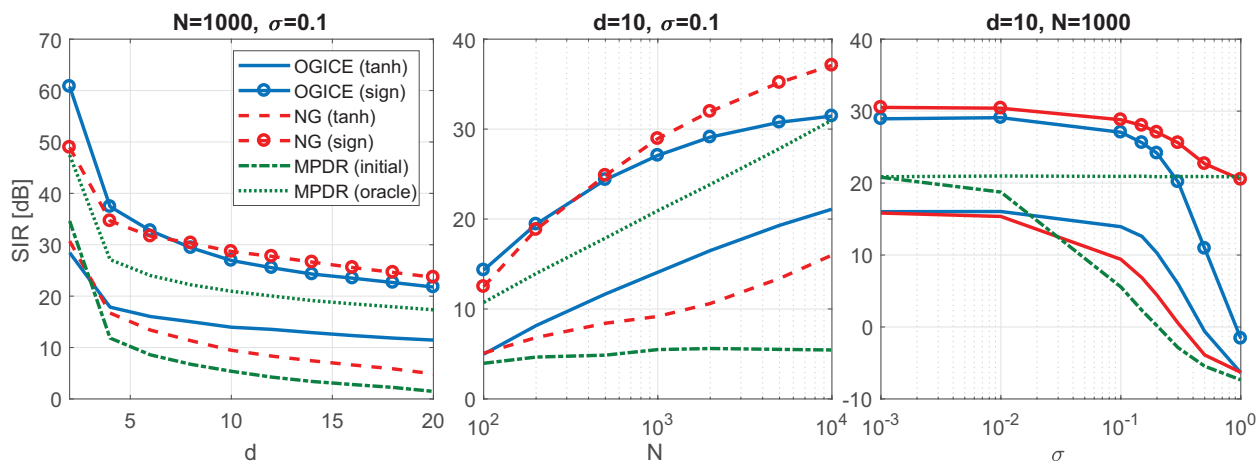
Fig. 2. Average Signal-to-Interference Ratio over 1000 trials, respectively, for varying d, N, and σ.

OGICE as well as by NG (with $\text{sign}(\cdot)$) points to the appealing potential of the blind estimators.

The dependence of SIR on N reveals that the accuracy of OGICE is limited when $N \to +\infty$. This is well known phenomenon caused by the orthogonal constraint [15], [20]. NG does not apply the constraint, so its accuracy grows faster with N. On the other hand, the results for $N < 10^3$ slightly promote OGICE prior to NG.

The results for varying σ reveal the radius of the region of convergence where the initial value leads to the separation of the desired signal (without random permutation in case of NG). Interestingly, NG appears to be robust in this example with $\text{sign}(\cdot)$ but less robust than OGICE with $\tanh(\cdot)$.

Last but not least, Fig. 1 shows the average computational time per one iteration as functions of d and N. In these examples, OGICE is approximately four times faster than NG (performed in Matlab on a PC with 4-core 2.7 GHz processor, 8 GB RAM). The algorithms need approximately the same number of iterations to converge.

V. CONCLUSIONS

The novel mixing model of ICE has been introduced, which contains the minimum number of parameters needed to handle the separation of one source of interest. A novel OG-constrained algorithm, OGICE, was proposed based on the mixing model. The method is simpler than NG as it does not aim to decompose the rest of the mixture into independent signals.

REFERENCES

[1] S. Makino, T.-W. Lee, and H. Sawada, Eds., *Blind Speech Separation*. Springer, 2007.
[2] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. John Wiley & Sons, Inc., 2002.
[3] O. Shalvi and E. Weinstein, "System identification using nonstationary signals," *IEEE Transactions on Signal Processing*, vol. 44, no. 8, pp. 2055–2063, Aug 1996.
[4] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.
[5] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*, ser. Independent Component Analysis and Applications Series. Elsevier Science, 2010.
[6] J. F. Cardoso, "Multidimensional independent component analysis," in *Proceedings of IEEE International Conference on Audio, Speech and Signal Processing*, vol. 4, May 1998, pp. 1941–1944 vol.4.
[7] S. Javidi, D. P. Mandic, and A. Cichocki, "Complex blind source extraction from noisy mixtures using second-order statistics," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1404–1416, July 2010.
[8] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Transactions on Audio, Speech, and Language Processing*, pp. 70–79, Jan. 2007.
[9] Y. O. Li, T. Adalı, W. Wang, and V. D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, Oct 2009.
[10] D. Lahat, T. Adalı, and C. Jutten, "Multimodal data fusion: An overview of methods, challenges, and prospects," *Proceedings of the IEEE*, vol. 103, no. 9, pp. 1449–1477, Sept 2015.
[11] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE Transactions on Speech and Audio Processing*, vol. 12, no. 5, pp. 530–538, Sep. 2004.
[12] J. F. Cardoso, "On extracting the cosmic microwave background from multi-channel measurements," in *Proceedings of International Conference on Latent Variable Analysis and Signal Separation*, Feb 2017, pp. 403–413.
[13] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Proceedings of Neural Information Processing Systems*, 1996, pp. 757–763.
[14] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug 2001.
[15] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proceedings of European Signal Processing Conference*, Sep. 1994, pp. 776–779.
[16] T. Adalı, M. Anderson, and G. S. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Processing Magazine*, vol. 31, no. 3, pp. 18–33, May 2014.
[17] J. F. Cardoso, "Blind signal separation: statistical principles," *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct 1998.
[18] A. Bell and T. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.
[19] J. F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Transactions on Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec 1996.
[20] Z. Koldovský, P. Tichavský, and E. Oja, "Efficient variant of algorithm FastICA for independent component analysis attaining the Cramér-Rao lower bound," *IEEE Transactions on Neural Networks*, vol. 17, no. 5, pp. 1265–1277, Sept 2006.