Informed Source Extraction based on Independent Vector Analysis using Eigenvalue Decomposition

Andreas Brendel and Walter Kellermann
Multimedia Communications and Signal Processing, Friedrich-Alexander-Universität Erlangen-Nürnberg,
Cauerstr. 7, D-91058 Erlangen, Germany, e-mail: Andreas.Brendel@FAU.de

***Abstract*—A desired acoustic source can very often only be observed in a mixture together with interfering sources in a real-life scenario. Hence, extracting the desired signal with a minimum amount of information about the geometric and acoustic setup is a problem of great interest. Recently, methods for blind source extraction based on Independent Vector Analysis (IVA) have been proposed. These algorithms are entirely blind, which prevents them from focussing on a specific source in the mixture. In this contribution, we guide the convergence of the extraction filter by a free-field prior within a Bayesian model towards the desired solution and use recently proposed update rules relying on the Eigenvalue Decomposition (EVD) for its optimization. The superiority of the presented update rules over a recently proposed state-of-the-art method is shown in experiments using measured Room Impulse Responses (RIRs).**

***Index Terms*—Independent Vector Analysis, Signal Extraction, Eigenvalue Decomposition, MM Algorithm**

I. Introduction

In daily-life situations, acoustic sources can often only be observed in mixtures with noise and interfering sources. Blind Source Separation (BSS) methods separate such observed mixtures to provide access to the individual sources of the mixture [1], [2]. Originating from instantaneous Independent Component Analysis (ICA) [3], Frequency-Domain ICA [4] has been proposed which solves the convolutive mixing problem, which is more relevant for acoustic mixtures, in the frequency domain. However, due to the independent treatment of the frequency bins, the inner permutation problem, i.e., an ambiguity of the permutation of the outputs in each frequency bin, occurs and has to be resolved afterwards [5]. To address this issue, Independent Vector Analysis (IVA) has been proposed [6], which avoids the inner permutation problem by imposing statistical dependencies between the frequency bins. Fast update rules based on the iterative projection principle and the Majorize-Minimize (MM) algorithm have been proposed [7], which can be considered as the gold standard in the field. For the special case of two sources and two microphones, even faster update rules have been developed based on the Generalized Eigenvalue Decomposition (GEVD) [8].

If BSS methods are used for extracting one or multiple Sources Of Interests (SOIs) from the observed mixture, additional information is needed in order to pick the desired source. A common way to achieve that is to incorporate spatial information in the adaptation of the SOI extraction filters [9]. This has been done for TRIple-N Independent component analysis for CONvolutive mixtures (TRINICON) in [10], [11]. For gradient-based IVA, spatial prior knowledge has been used in [12] and for iterative projection-based IVA in [13]. Another approach for SOI selection is to incorporate a reference signal, which is related to the SOI, into the SOI model [14], [15].

If the number of SOIs is smaller than the number of microphones, a Background (BG) model can be used to reduce the number of parameters to be estimated and hence to reduce the computational complexity of the algorithm [16]. For IVA using update rules based on iterative projection, such a BG model has been introduced by [17] and fast update rules have been proposed for this model by [18]. In [19], a generic Bayesian framework has been developed which contains a BG model and allows for many ways to incorporate spatial and other prior knowledge about the source into the adaptation of the SOI extraction filters.

In this contribution, we aim at extracting a single SOI from a mixture of recorded sources. To this end, we extend recently proposed update rules based on Eigenvalue Decomposition (EVD) [18] for IVA-based source extraction by a generic way to incorporate prior knowledge about the SOI extraction filter. In [18], no such SOI selection mechanism is included and hence always the dominant source of a mixture is taken as the SOI. However, in many realistic situations multiple sources will be similarly dominant as the SOI and hence the extraction will not necessarily yield the desired result. The paper presented here provides an extension of the framework presented in [19] and the newly introduced update rules lead to fast convergence at a low computational cost. In experiments using recorded Room Impulse Responses (RIRs), it has been shown that the proposed updates are superior to recently proposed state-of-the-art methods [19].

The remainder of the paper is structured as follows: Sec. II introduces the signal models and the probabilistic model, which allows to incorporate prior knowledge about the SOI extraction filters. Update rules are derived in Sec. III. Results are presented and discussed in Sec. IV, and the paper is concluded by Sec. V.

II. Model

In this section, we describe the signal model and the probabilistic model which leads to the optimization problem treated in Sec. III.

This work was supported by DFG under contract no <Ke890/10-1> within the Research Unit FOR2457 "Acoustic Sensor Networks".

A. Signal Model

We consider a linear array consisting of M microphones, observing an acoustic scene containing one SOI among multiple interferers. The Direction of Arrival (DOA) of the SOI is assumed to be known.

The output Short-Time Fourier Transform (STFT)-domain signals $\mathbf{y}_{f,n}$ at time frame $n \in \{1, \ldots, N\}$ and frequency subband $f \in \{1, \ldots, F\}$ is obtained by demixing the observed microphone signals

$$\mathbf{x}_{f,n} = [x_{1,f,n}, \ldots, x_{M,f,n}]^{\mathrm{T}} \in \mathbb{C}^M \tag{1}$$

by the demixing matrix $\mathbf{W}_f$

$$\mathbf{y}_{f,n} = \begin{bmatrix} s_{f,n} \\ \mathbf{z}_{f,n} \end{bmatrix} = \mathbf{W}_f \mathbf{x}_{f,n}. \tag{2}$$

The output signal vector contains the SOI $s_{f,n}$ as well as a set of so-called BG signals

$$\mathbf{z}_{f,n} = [z_{1,f,n}, \ldots, z_{M-1,f,n}]^{\mathrm{T}} \in \mathbb{C}^{M-1}, \tag{3}$$

for which we will not spend any efforts to separate them. The demixing matrix

$$\mathbf{W}_f = \begin{bmatrix} \mathbf{w}_f^{\mathrm{H}} \\ \mathbf{B}_f \end{bmatrix} \in \mathbb{C}^{M \times M} \tag{4}$$

consists of two parts: the SOI filter $\mathbf{w}_f \in \mathbb{C}^M$ and the BG filter matrix $\mathbf{B}_f \in \mathbb{C}^{(M-1)\times M}$. The broadband SOI vector is defined as

$$\underline{\mathbf{s}}_n = [s_{f,n}, \ldots, s_{F,n}]^{\mathrm{T}} \in \mathbb{C}^F \tag{5}$$

and the broadband vector containing the BG signals is defined as

$$\underline{\mathbf{z}}_n = \left[\underline{\mathbf{z}}_{1,n}^{\mathrm{T}}, \ldots, \underline{\mathbf{z}}_{M-K,n}^{\mathrm{T}}\right]^{\mathrm{T}} \in \mathbb{C}^{(M-1)F}. \tag{6}$$

B. Probabilistic Model

We denote the set of demixing matrices defined in (4) by $\mathcal{W}$ and the set of demixed signal vectors and microphone signals of all time frames and frequency bins by $\mathcal{Y}$ and $\mathcal{X}$, respectively. Using these definitions, we can use Bayes' Theorem to obtain an expression of the posterior distribution of the demixing matrices and demixed signals

$$p(\mathcal{W}, \mathcal{Y}|\mathcal{X}) \propto p(\mathcal{W}) p(\mathcal{Y}|\mathcal{W}) p(\mathcal{X}|\mathcal{W}, \mathcal{Y}). \tag{7}$$

Under an i.i.d. assumption over time frames n and frequency bins f, we construct the following likelihood function

$$p(\mathcal{X}|\mathcal{W}, \mathcal{Y}) = \prod_{n=1}^{N} \prod_{f=1}^{F} \delta\left(\mathbf{x}_{f,n} - \mathbf{W}_f^{-1} \mathbf{y}_{f,n}\right), \tag{8}$$

where $\delta(\cdot)$ denotes the Dirac distribution. For the distribution of the demixed signals, we assume independence between all time frames

$$p(\mathcal{Y}|\mathcal{W}) = \prod_{n=1}^{N} p\left(\underline{\mathbf{y}}_n\right) = \prod_{n=1}^{N} p\left(\underline{\mathbf{z}}_n\right) p\left(\underline{\mathbf{s}}_n\right). \tag{9}$$

Hereby, the rightmost term expresses the assumption about mutual independence between the BG signals and the SOI. The posterior density of the demixing matrices is obtained by marginalization of (7) over the demixed signals $\mathcal{Y}$

$$p(\mathcal{W}|\mathcal{X}) \propto p(\mathcal{W}) \int p(\mathcal{Y}|\mathcal{W}) p(\mathcal{X}|\mathcal{W}, \mathcal{Y}) d\underline{\mathbf{y}}_1 \ldots d\underline{\mathbf{y}}_N, \tag{10}$$

which yields after inserting the models (8) and (9)

$$p(\mathcal{W}|\mathcal{X}) \propto p(\mathcal{W}) \prod_{n=1}^{N} \int p(\underline{\mathbf{y}}_n) \prod_{f=1}^{F} \delta\left(\mathbf{x}_{f,n} - \mathbf{W}_f^{-1} \mathbf{y}_{f,n}\right) d\underline{\mathbf{y}}_n. \tag{11}$$

By using the rules for a linear transform of complex variables [20] to the transform $\mathbf{y}_{f,n} = \mathbf{W}_f \mathbf{x}_{f,n}$ and the sifting property of the Dirac distribution we obtain

$$p(\mathcal{W}|\mathcal{X}) \propto p(\mathcal{W}) \prod_{f=1}^{F} |\det \mathbf{W}_f|^{2N} \prod_{n=1}^{N} p\left(\underline{\mathbf{z}}_n\right) p(\underline{\mathbf{s}}_n). \tag{12}$$

By taking the logarithm of (12), we finally obtain the following Maximum A Posteriori (MAP) problem

$$\begin{aligned} \mathbf{w}_f = \underset{\mathbf{w}_f \in \mathbb{C}^M}{\arg\max}\, & \frac{\log p(\mathcal{W})}{N} + 2 \sum_{f=1}^{F} \log|\det \mathbf{W}_f| \ldots \\ & \cdots - \hat{\mathbb{E}}\left\{G\left(\underline{\mathbf{s}}_n\right)\right\} + \hat{\mathbb{E}}\left\{\log p\left(\underline{\mathbf{z}}_n\right)\right\}, \end{aligned} \tag{13}$$

where we defined the score function $G(\underline{\mathbf{s}}_n) = -\log p(\underline{\mathbf{s}}_n)$ and the averaging operator $\hat{\mathbb{E}}\{\cdot\} = \frac{1}{N}\sum_{n=1}^{N}(\cdot)$ for a concise notation.

By taking the negative of the MAP problem (13), we obtain the following minimization problem

$$\mathbf{w}_f = \underset{\mathbf{w}_f \in \mathbb{C}^M}{\arg\min}\, J(\mathbf{w}_f), \tag{14}$$

which will be used in the following. In the derivation of the update rules, we will apply a prewhitening step to simplify the update rules in Sec. III-C, which is accomplished by multiplication with the matrix $\mathbf{Q}_f^{-\mathrm{H}}$ obtained from the Cholesky decomposition of the microphone correlation matrix

$$\mathbf{C}_f = \hat{\mathbb{E}}\left\{\mathbf{x}_{f,n} \mathbf{x}_{f,n}^{\mathrm{H}}\right\} = \mathbf{Q}_f^{\mathrm{H}} \mathbf{Q}_f. \tag{15}$$

C. SOI Distributions

For modeling the source Probability Density Functions (PDFs), we use two popular models [21]: The generalized Gaussian distribution

$$p\left(\underline{\mathbf{s}}_n\right) \propto \exp\left(-\|\underline{\mathbf{s}}_n\|_2^{\beta}\right), \tag{16}$$

with the shape parameter $\beta \in \mathbb{R}_+$, and the time-varying Gaussian distribution

$$p\left(\underline{\mathbf{s}}_n\right) \propto \exp\left(-\frac{\|\underline{\mathbf{s}}_n\|_2^2}{\sigma_n^2}\right). \tag{17}$$

D. Background Model

The BG signals are modeled to be normally distributed

$$p\left(\mathbf{z}_{f,n}\right) = \frac{1}{\pi^{M-K}|\det \mathbf{R}_f|} \exp\left(-\mathbf{z}_{f,n}^{\mathrm{H}} \mathbf{R}_f^{-1} \mathbf{z}_{f,n}\right) \tag{18}$$

and independent over all time frames n and frequency bins f

$$p(\underline{\mathbf{z}}_1, \ldots, \underline{\mathbf{z}}_N) = \prod_{n=1}^{N} p\left(\underline{\mathbf{z}}_n\right) = \prod_{n=1}^{N} \prod_{f=1}^{F} p\left(\mathbf{z}_{f,n}\right). \tag{19}$$

E. Priors

To incorporate prior knowledge, we have to choose an expression for the prior term $p(\mathcal{W})$. First, we assume the priors of the demixing matrices to be independent over all frequency bins and the prior of the SOI filter to be independent from the prior of the BG filters

$$p(\mathcal{W}) = \prod_{f=1}^{F} p(\mathbf{W}_f) = \prod_{f=1}^{F} p\left(\mathbf{B}_f\right) p\left(\mathbf{w}_f\right). \tag{20}$$

As we do not want to identify the BG signals, we also do not incorporate prior knowledge about the corresponding filters and choose an uninformative prior for $p\left(\mathbf{B}_f\right)$. For the prior of the SOI filter, we choose the following Gaussian distribution with precision matrix $\tilde{\gamma}_f \mathbf{P}_f$ [19]

$$p\left(\mathbf{w}_f\right) = \frac{\sqrt{(\tilde{\gamma}_f)^M \det \mathbf{P}_f}}{\sqrt{\pi^M}} \exp\left(-\tilde{\gamma}_f (\mathbf{w}_f)^{\mathrm{H}} \mathbf{P}_f \mathbf{w}_f\right). \tag{21}$$

The precision matrices are chosen according to

$$\mathbf{P}_f = \alpha_{\mathrm{Tik}} \mathbf{I}_M - \alpha \mathbf{Q}_f^{\mathrm{H}} \mathbf{h}_f(\vartheta) \mathbf{h}_f(\vartheta)^{\mathrm{H}} \mathbf{Q}_f, \tag{22}$$

where the first part corresponds to a Tikhonov regularization, i.e., a constraint on the filters energy, and the second part forces the filter to be parallel to the prototype filters $\mathbf{h}_f(\vartheta)$ multiplied with a unitary whitening matrix $\mathbf{Q}_f^{\mathrm{H}}$ (the reason for this operation will become clear in Sec. III-C). The parameters α_{Tik} and α weight the influence of the different components of the precision matrix. As an example for applying prior spatial knowledge, we choose here the prototype filters $\mathbf{h}_f(\vartheta)$ to represent relative delays in a free-field model

$$[\mathbf{h}_f(\vartheta)]_m = \left[\exp\left(j \frac{2\pi\mu_f}{c_s} \|\mathbf{r}_m - \mathbf{r}_1\|_2 \cos\vartheta\right)\right]_m, \tag{23}$$

where m is the microphone index, μ_f the frequency corresponding to frequency bin f in Hertz, c_s the speed of sound, $\mathbf{r}_m$ the position of the mth microphone and ϑ the DOAs, to which we want to steer a spatial one.

III. Optimization

Representing the essence of this contribution, we construct in the following a surrogate for the optimization problem (14) in the form of the upper bound (33), which will allow for the derivation of fast and computationally efficient update rules.

A. MM Principle

The main idea behind MM algorithms [22] is to construct an upper bound $U\left(\mathcal{W}|\mathcal{W}^{(l)}\right)$ of the cost function

$$J(\mathcal{W}) \leq U\left(\mathcal{W}|\mathcal{W}^{(l)}\right), \tag{24}$$

where $\mathcal{W}^{(l)}$ denotes the set of all SOI filters estimated in iteration $l \in \{1, \ldots, L\}$ and equality holds iff $\mathcal{W} = \mathcal{W}^{(l)}$, i.e.,

$$J\left(\mathcal{W}^{(l)}\right) = U\left(\mathcal{W}^{(l)}|\mathcal{W}^{(l)}\right). \tag{25}$$

This upper bound should be easier to optimize than the cost function itself, such that the SOI filters of the next iteration can be obtained by

$$\mathcal{W}^{(l+1)} = \underset{\mathcal{W}}{\operatorname{argmin}}\, U\left(\mathcal{W}|\mathcal{W}^{(l)}\right). \tag{26}$$

Together, this yields the following inequality, which assures the cost function to be non-increasing due to the updates of the MM algorithm

$$\begin{aligned} J\left(\mathcal{W}^{(l+1)}\right) &\leq U\left(\mathcal{W}^{(l+1)}|\mathcal{W}^{(l)}\right) \\ &\leq U\left(\mathcal{W}^{(l)}|\mathcal{W}^{(l)}\right) = J\left(\mathcal{W}^{(l)}\right). \end{aligned} \tag{27}$$

B. Upper Bound

The following inequality has been proven in [7] for IVA

$$\hat{\mathbb{E}}\left\{G(\underline{\mathbf{s}}_n)\right\} \leq R(\mathcal{W}^{(l)}) + \frac{1}{2} \sum_{f=1}^{F} (\mathbf{w}_f)^{\mathrm{H}} \mathbf{V}_f \left(\mathcal{W}^{(l)}\right) \mathbf{w}_f. \tag{28}$$

Hereby, $R\left(\mathcal{W}^{(l)}\right)$ denotes a term independent of $\mathbf{w}_f$. Note that for the time-varying Gaussian distribution (17) equality holds in (28) and $R(\mathcal{W}^{(l)}) = 0$. The matrix $\mathbf{V}_f\left(\mathcal{W}^{(l)}\right)$ denotes a weighted microphone signals' covariance matrix

$$\mathbf{V}_f\left(\mathcal{W}^{(l)}\right) = \hat{\mathbb{E}}\left\{\phi(r_n) \tilde{\mathbf{x}}_{f,n} \tilde{\mathbf{x}}_{f,n}^{\mathrm{H}}\right\}, \tag{29}$$

where the weighting factor is defined by

$$\phi(r_n) = \frac{\tilde{G}'\left(r_n\left(\mathcal{W}^{(l)}\right)\right)}{r_n\left(\mathcal{W}^{(l)}\right)}. \tag{30}$$

The broadband signal standard deviation $r_n\left(\mathcal{W}^{(l)}\right)$

$$r_n\left(\mathcal{W}^{(l)}\right) = \left\|\underline{\mathbf{s}}_n^{(l)}\right\|_2 = \sqrt{\sum_{f=1}^{F} \left|\left(\mathbf{w}_f^{(l)}\right)^{\mathrm{H}} \mathbf{x}_{f,n}\right|^2} \tag{31}$$

can also be used to write the SOI model of Sec. II-C as $G(\underline{\mathbf{s}}_n) = \tilde{G}(r_n(\mathcal{W}_k^{(l)}))$. For the presented SOI models, the weighting factor $\phi(r_n)$ of the weighted covariance matrix (29) can in a generalized form be expressed (see [21])

$$\phi(r_n) = (r_n)^{\beta-2}. \tag{32}$$

By applying the inequality (28) to the cost function (14), we obtain the following upper bound for the cost function J

$$\begin{aligned} U\left(\mathcal{W}|\mathcal{W}^{(l)}\right) = \sum_{f=1}^{F} \Bigg[&\frac{1}{2} \mathbf{w}_f^{\mathrm{H}} \mathbf{V}_f \left(\mathcal{W}^{(l)}\right) \mathbf{w}_f \ldots \\ &\ldots + \frac{1}{F} R\left(\mathcal{W}^{(l)}\right) - 2\log|\det \mathbf{W}_f| + \gamma_f \mathbf{w}_f^{\mathrm{H}} \mathbf{P}_f \mathbf{w}_f \Bigg], \end{aligned} \tag{33}$$

where $\gamma_f = \frac{\tilde{\gamma}_f}{N}$.

C. Optimization of Upper Bound

Optimization of the upper bound (33) w.r.t. $\mathbf{W}_f$ yields the following conditions for the SOI and BG filters

$$\mathbf{w}_f^{\mathrm{H}} \left[\mathbf{V}_f^k\left(\mathcal{W}^{(l)}\right) + \gamma_f \mathbf{P}_f\right] \mathbf{w}_f \stackrel{!}{=} 1, \tag{34}$$

$$\mathbf{B}_f^{\mathrm{H}} \left[\mathbf{V}_f^k\left(\mathcal{W}^{(l)}\right) + \gamma_f \mathbf{P}_f\right] \mathbf{w}_f \stackrel{!}{=} \mathbf{0}, \tag{35}$$

$$\mathbf{B}_f^{\mathrm{H}} \mathbf{C}_f \mathbf{B}_f \stackrel{!}{=} \mathbf{R}_f, \tag{36}$$

$$\mathbf{w}_f^{\mathrm{H}} \mathbf{C}_f \mathbf{B}_f \stackrel{!}{=} \mathbf{0}. \tag{37}$$

Similarly to [18] the optimal $\mathbf{w}_f$ can be obtained as

$$\mathbf{w}_f = \frac{1}{\sqrt{\lambda_{M,f}}} \mathbf{Q}_f^{-1} \mathbf{u}_{M,f}, \tag{38}$$

where λ_M and $\mathbf{u}_M$ denotes the smallest eigenvalue and eigenvector, respectively, of

$$\mathbf{V}_f^{\mathrm{inf}} = \mathbf{Q}_f^{-\mathrm{H}} \left[\mathbf{V}_f^k\left(\mathcal{W}^{(l)}\right) + \gamma_f \mathbf{P}_f\right] \mathbf{Q}_f^{-1}. \tag{39}$$

Note that the incorporation of the prior knowledge, which allows for selection of the SOI, causes the additional term $\gamma_f \mathbf{P}_f$ in (39) compared to the updates presented in [18], which will select the dominant source as SOI. To iteratively minimize the cost function corresponding to (13), we have to alternate between two steps: Minimization of the upper bound (33) by (39) and construction of a new upper bound by computing the weighted microphone covariance matrix (29) and broadband signal standard deviation (31).

The resulting update rules are summarized in Algorithm 1.

Algorithm 1 Guided Signal Extraction

INPUT: $\mathcal{X}$, L, ϑ

INITIALIZATION:
$s_{f,n} = x_{1,f,n}\ \forall f, n$
$\mathbf{W}_f^{(0)} = \mathbf{I}_M\ \forall f$
Whitening $\tilde{\mathbf{x}}_{f,n} = \mathbf{Q}_f^{-\mathrm{H}} \mathbf{x}_{f,n}$

for $l = 1$ **to** L **do**
  Calculate $\phi(r_n)\ \forall n$
  for $f = 1$ **to** F **do**
    Calculate $\tilde{\mathbf{V}}_f(\mathcal{W}_k^{(l)}) = \hat{\mathbb{E}}\left\{\phi(r_n)\tilde{\mathbf{x}}_{f,n}\tilde{\mathbf{x}}_{f,n}^{\mathrm{H}}\right\}$
    Compute smallest eigenvalue $\lambda_{M,f}$ and eigenvector $\mathbf{u}_{M,f}$ of $\mathbf{V}_f^{\mathrm{inf}} = \gamma_f \mathbf{Q}_f^{-\mathrm{H}} \mathbf{P}_f \mathbf{Q}_f^{-1} + \tilde{\mathbf{V}}_f(\mathcal{W}_k^{(l)})$
    $\mathbf{w}_f = \frac{1}{\sqrt{\lambda_{M,f}}} \mathbf{u}_{M,f}$
    $\mathbf{w}_f \leftarrow \frac{\mathbf{w}_f}{\|\mathbf{w}_f\|_2}$
    Extract SOI $s_{f,n} = \mathbf{w}_f^{\mathrm{H}} \tilde{\mathbf{x}}_{f,n}\ \forall n$
  end for
end for

OUTPUT: SOI $s_{f,n}\, \forall f, n$

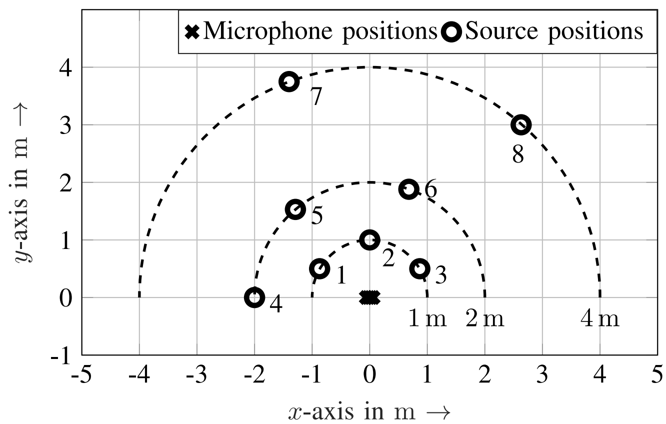

Fig. 1. Position of the sources and microphones for the measured RIRs. The position of the array containing $M = 4$ microphones is depicted by crosses and the 8 source positions are marked by circles.

IV. Experiments

In the following, we will present results of the proposed update rules and benchmark them with a recently proposed baseline using updates based on iterative projection, i.e., Algorithm 3 in [19]. In this reference, a comprehensive comparative study with other state-of-the-art methods can also be found for the same experimental setup.

For the experiments, we measured RIRs of 8 source positions at $1\,\mathrm{m}$, $2\,\mathrm{m}$ and $4\,\mathrm{m}$ distance from the center of the microphone array at the same height of $1.4\,\mathrm{m}$ in an office room with a reverberation time of $T_{60} \approx 0.2\,\mathrm{s}$. The setup is depicted in Fig. 1. The sources, represented by 4 female and 4 male speech signals, are observed by a linear microphone array containing $M = 4$ microphones with a spacing of $4.2\,\mathrm{cm}$. White Gaussian noise is added to the microphone signals to simulate sensor noise at a Signal-to-Noise Ratio (SNR) of $30\,\mathrm{dB}$. The microphone signals are transformed to the STFT domain using a von Hann window of 2048 samples length at a sampling frequency of $16\,\mathrm{kHz}$ and 50% overlap. The performance of the algorithms is measured in terms of the improvement of Signal-to-Distortion Ratio (SDR), Signal-to-Interference Ratio (SIR) and Signal-to-Artefact Ratio (SAR) [23], respectively, relative to the input signals. For representative results, we changed the association of source signals to positions 20 times and averaged over the results. For the parameters of the prior (21) and the source model (32), we chose $\beta = 0$, $\alpha_{\text{Tik}} = 0.01$, $\alpha = 0.0005$ for the proposed method, and $\beta = 1$, $\alpha_{\text{Tik}} = 1.5$, $\alpha = 3$ for the baseline method [19]. Note that $\beta = 1$ corresponds to a Laplacian SOI model and $\beta = 0$ to a time-varying Gaussian SOI model (17). The parameters are chosen w.r.t. the extraction of Source 2 (see Fig. 1) such that the extraction of the desired source is assured and the best performance measures are achieved on average over all 20 trials.

The results for the extraction of sources 1, 2 and 3 are depicted in Fig. 2 in terms of SDR, SIR and SAR improvement over the number of iterations. It can be seen that the proposed

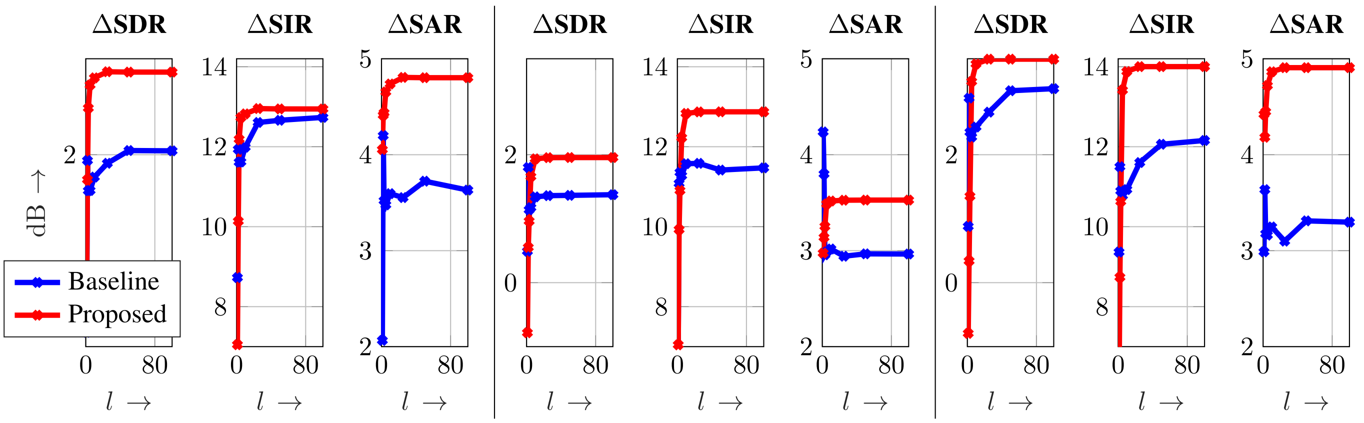

Fig. 2. Performance of the extraction of Source 1, 2 and 3 (see Fig. 1) in terms of SDR, SIR and SAR improvement over the number of iterations. The proposed method is plotted in red, the baseline method [19] is plotted in blue.

updates provide faster convergence and better final results than the baseline methods. The proposed method reaches convergence after $2-5$ iterations. In some cases, a non-monotonous behavior can be seen in the curves of Fig. 2. However, this is not a contradiction as the optimization of the performance measures is not the objective of the cost function (14). In terms of runtime, the proposed method and the baseline are in the same order of magnitude, with a significant advantage for the proposed method: The average runtime per iteration is about 0.25 seconds for the baseline and 0.08 seconds for the proposed updates using the EVD provided by Matlab.

V. Conclusions

In this paper, we proposed fast converging and computationally simple update rules for informed source extraction based on IVA. The proposed updates are shown to provide faster convergence at a lower runtime than a recently proposed baseline method. The proposed updates are an extension of the update rules proposed by [18], where no SOI selection mechanism is used. Furthermore, the proposed updates represent an extension to the framework [19], where no such updates have been considered.

References

[1] S. Makino, T.-W. Lee, and H. Sawada, Eds., *Blind speech separation*, ser. Signals and communication technology. Dordrecht: Springer, 2007.

[2] E. Vincent, T. Virtanen, and S. Gannot, Eds., *Audio source separation and speech enhancement*. Hoboken, NJ: John Wiley & Sons, 2018.

[3] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. New York: J. Wiley, 2001.

[4] P. Smaragdis, "Blind Separation of Convolved Mixtures in the Frequency Domain," *Neurocomputing J.*, vol. 22, pp. 21–34, 1998.

[5] 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 Trans. on Speech and Audio Process.*, vol. 12, no. 5, pp. 530–538, Sep. 2004.

[6] T. Kim et al., "Blind Source Separation Exploiting Higher-Order Frequency Dependencies," *IEEE Trans. on Audio, Speech and Language Process.*, vol. 15, no. 1, pp. 70–79, Jan. 2007.

[7] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *IEEE Workshop on Applicat. of Signal Process. to Audio and Acoust. (WASPAA)*, New Paltz, NY, USA, Oct. 2011, pp. 189–192.

[8] ——, "Fast Stereo Independent Vector Analysis and its Implementation on Mobile Phone," in *Int. Workshop on Acoustic Signal Enhancement (IWAENC)*, Aachen, Germany, Sep. 2012.

[9] L. Parra and C. Alvino, "Geometric source separation: merging convolutive source separation with geometric beamforming," *IEEE Trans. on Speech and Audio Process.*, vol. 10, no. 6, pp. 352–362, Sep. 2002.

[10] K. Reindl, S. Meier, H. Barfuss, and W. Kellermann, "Minimum Mutual Information-Based Linearly Constrained Broadband Signal Extraction," *IEEE/ACM Trans. on Audio, Speech, and Language Process.*, vol. 22, no. 6, pp. 1096–1108, Jun. 2014.

[11] Y. Zheng, K. Reindl, and W. Kellermann, "Analysis of dual-channel ICA-based blocking matrix for improved noise estimation," *EURASIP J. on Advances in Signal Process.*, vol. 2014, no. 1, Dec. 2014.

[12] A. H. Khan, M. Taseska, and E. A. P. Habets, "A Geometrically Constrained Independent Vector Analysis Algorithm for Online Source Extraction," in *Latent Variable Analysis and Signal Separation*. Cham: Springer Int. Publishing, 2015, vol. 9237, pp. 396–403.

[13] A. Brendel, T. Haubner, and W. Kellermann, "Spatially Guided Independent Vector Analysis," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Barcelona, Spain, May 2020.

[14] T. Ono, N. Ono, and S. Sagayama, "User-guided independent vector analysis with source activity tuning," in *IEEE Int. Conf. on Acoustic, Speech and Signal Process. (ICASSP)*, Kyoto, Japan, Mar. 2012, pp. 2417–2420.

[15] F. Nesta and Z. Koldovský, "Supervised independent vector analysis through pilot dependent components," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Mar. 2017, pp. 536–540.

[16] Z. Koldovský and P. Tichavský, "Gradient Algorithms for Complex Non-Gaussian Independent Component/Vector Extraction, Question of Convergence," *IEEE Trans. on Signal Process.*, vol. 67, no. 4, pp. 1050–1064, Feb. 2019.

[17] R. Scheibler and N. Ono, "Independent Vector Analysis with More Microphones than Sources," in *IEEE Workshop on Applications of Signal Process. to Audio and Acoust. (WASPAA)*, Oct. 2019.

[18] ——, "Fast Independent Vector Extraction by Iterative SINR Maximization," in *IEEE Int. Conf. on Acoust., Speech and Signal Process. (ICASSP)*, Barcelona, Spain, May 2020, pp. 601–605.

[19] A. Brendel, T. Haubner, and W. Kellermann, "A Unified Probabilistic View on Spatially Informed Source Separation and Extraction based on Independent Vector Analysis," *accepted for: IEEE Trans. on Signal Process.*, 2020.

[20] E. Moreau and T. Adali, *Blind identification and separation of complex-valued signals*, ser. Focus series in digital signal and image processing. London: Hoboken, NJ: ISTE; Wiley, 2013.

[21] N. Ono, "Auxiliary-function-based independent vector analysis with power of vector-norm type weighting functions," in *Asia-Pacific Signal & Inform. Process. Association Annual Summit and Conf. (APSIPA ASC)*, Dec. 2012.

[22] D. R. Hunter and K. Lange, "A Tutorial on MM Algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[23] E. Vincent, R. Gribonval, and C. Févotte, "Performance measurement in blind audio source separation," *IEEE Trans. on Audio, Speech and Language Process.*, vol. 14, no. 4, pp. 1462–1469, Jul. 2006.