Auxiliary-Function-Based Steering Vector Estimation Method for Spatially Regularized Independent Low-Rank Matrix Analysis

Sota Hirata*, Norihiro Takamune*, Kouei Yamaoka*, Daichi Kitamura†,
Hiroshi Saruwatari*, Yu Takahashi‡, and Kazunobu Kondo‡
* The University of Tokyo, Tokyo, Japan
† National Institute of Technology, Kagawa College, Kagawa, Japan
‡ Yamaha Corporation, Shizuoka, Japan

***Abstract*—Blind source separation (BSS) is a technique to separate each source signal from observed mixtures without prior information, and independent low-rank matrix analysis (ILRMA) is one of the state-of-the-art BSS methods. As not a fully blind method, spatially regularized ILRMA (SR-ILRMA) achieves higher separation performance by using prior information about the demixing matrix or steering vectors (SVs) as a regularizer in ILRMA. In this paper, we propose a new fully blind method that models SVs using the time differences of arrival (TDOAs) and simultaneously estimates TDOAs, the demixing matrix, and other parameters. Then, we derive the update rule of TDOAs on the basis of the majorization-minimization algorithm, which guarantees the monotonic non-increase in the proposed cost function. We also propose the method of initializing TDOAs based on impulse responses for the proposed cost function. Numerical experiments confirmed that the proposed method achieves better separation performance than ILRMA, and initialization based on impulse responses is effective for the proposed method.**

I. Introduction

Blind source separation (BSS) [1] is a technique to separate each source signal from mixtures observed by a microphone array without any information about the transmission system or the characteristics of the source. BSS is an essential preprocessing technique for acoustic signal processing in real-world applications, such as speech recognition and hearing aids. Representative methods of BSS include frequency-domain independent component analysis (FDICA) [2], independent vector analysis (IVA) [3], [4], and independent low-rank matrix analysis (ILRMA) [5]. FDICA estimates the demixing matrix by assuming statistical independence between sources for each frequency bin. However, the outputs in each frequency bin are unordered, and thus, FDICA should align outputs over all frequency bins (permutation problem). IVA and ILRMA simultaneously estimate the demixing matrix and solve the permutation problem using the generative model that has the higher-order correlation between frequency bins in each source. In particular, ILRMA assumes a more sophisticated model than IVA and experimentally achieves a high separation performance [5]. However, it has been reported that there is still room for improvement in the accuracy of solving the permutation problem even in ILRMA [6].

In contrast to fully blind methods, several methods use spatial prior information about the source directions and the microphone array geometry to improve the separation performance [7]–[9]. This spatial prior information often includes some errors due to measurement errors in the directions of the sources and the position of the microphone array, microphone directivity, diffraction around the microphone array, and reverberation, among others. Therefore, in [7]–[9], such prior information is utilized as a regularizer to tolerate some errors. Spatially regularized ILRMA (SR-ILRMA) [9] applies such regularizer to ILRMA and it has been reported that SR-ILRMA achieves higher separation performance than conventional ILRMA.

In terms of applicability, a fully blind method is desirable. In this paper, we propose a new fully blind method that estimates the time differences of arrival (TDOAs), which implicitly retain information about the source directions and the microphone array geometry, simultaneously with the demixing matrix. In [7], the demixing matrix and source directions were simultaneously estimated using different criteria for each. In contrast, we aim to formulate this simultaneous estimation problem as a single minimization problem. Then, for the majorization-minimization (MM) algorithm [10], we design the auxiliary function of the proposed cost function with respect to TDOAs and derive the update rule that guarantees the monotonic non-increase in the cost function. We also investigated the method of initializing TDOAs for the proposed method.

First, we investigated the initial value dependence of TDOAs in the preliminary experiment. Then, we conducted a numerical experiment to confirm the effectiveness of the proposed method.

II. Conventional methods

A. MM algorithm [10]

The MM algorithm is an optimization method used to find $\boldsymbol{s}$ that minimizes the cost function $f(\boldsymbol{s})$. For a given cost function $f(\boldsymbol{s})$, the auxiliary function $f^{+}(\boldsymbol{s}, \tilde{\boldsymbol{s}})$ is designed to satisfy the

This work was supported by JST Moonshot R&D Grant Number JPMJMS2011 (for algorithm design) and Tateisi Science and Technology Foundation (for numerical experiment).

following conditions:

$$f(\boldsymbol{s}) \leq f^{+}(\boldsymbol{s}, \tilde{\boldsymbol{s}}), \quad \forall \boldsymbol{s}, \tilde{\boldsymbol{s}}, \tag{1}$$

$$f(\tilde{\boldsymbol{s}}) = f^{+}(\tilde{\boldsymbol{s}}, \tilde{\boldsymbol{s}}), \quad \forall \tilde{\boldsymbol{s}}, \tag{2}$$

where $\tilde{\boldsymbol{s}}$ is an auxiliary variable. When the auxiliary function $f^{+}(\boldsymbol{s}, \tilde{\boldsymbol{s}})$ exists, the update rules for the MM algorithm are

$$\hat{\boldsymbol{s}} \leftarrow \boldsymbol{s}, \tag{3}$$

$$\boldsymbol{s} \leftarrow \underset{\boldsymbol{s}}{\operatorname{argmin}}\ f^{+}(\boldsymbol{s}, \hat{\boldsymbol{s}}). \tag{4}$$

The cost function $f(\boldsymbol{s})$ is minimized by the iterative update of (3) and (4) from a given initial value of $\boldsymbol{s}$.

The MM algorithm guarantees the monotonic non-increase in the cost function as

$$f(\boldsymbol{s}^{(l)}) = f^{+}(\boldsymbol{s}^{(l)}, \boldsymbol{s}^{(l)}) \geq f^{+}(\boldsymbol{s}^{(l+1)}, \boldsymbol{s}^{(l)}) \geq f(\boldsymbol{s}^{(l+1)}), \tag{5}$$

where the superscript $^{(l)}$ denotes the parameter updated in the lth iteration. Unlike other gradient-based optimization methods, the MM algorithm does not require step size tuning. The most crucial aspect of the MM algorithm is designing an auxiliary function for each cost function of the optimization problem. Auxiliary functions are not unique for each cost function; thus, it is important to find a better auxiliary function for achieving faster convergence.

B. ILRMA [5]

Let $\boldsymbol{x}_{ij} = (x_{ij1}, ..., x_{ijM})^{\mathsf{T}} \in \mathbb{C}^{M}$, $\boldsymbol{s}_{ij} = (s_{ij1}, ..., s_{ijN})^{\mathsf{T}} \in \mathbb{C}^{N}$, and $\boldsymbol{y}_{ij} = (y_{ij1}, ..., y_{ijN})^{\mathsf{T}} \in \mathbb{C}^{N}$ be the short-time Fourier transforms (STFTs) of the observed, source, and separated signals, respectively. Here, $i \in \{1, ..., I\}$, $j \in \{1, ..., J\}$, $m \in \{1, ..., M\}$, and $n \in \{1, ..., N\}$ are the indices of the frequency bins, time frames, microphones, and sources, respectively, and $\cdot^{\mathsf{T}}$ denotes the transpose. If each source is a point source and the reverberation time is sufficiently shorter than the window length of the STFT, the observed signal is approximately represented as $\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij}$. Here, $\boldsymbol{A}_i = (\boldsymbol{a}_{i1}, ...\boldsymbol{a}_{iN}) \in \mathbb{C}^{M \times N}$ is the mixing matrix representing the time-invariant spatial characteristics of the transmission system, and $\boldsymbol{a}_{in}$ is the steering vector (SV) of the nth source. If $M = N$ and $\boldsymbol{A}_i$ is regular, the separated signals can be obtained as

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}, \tag{6}$$

where $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, ..., \boldsymbol{w}_{iN})^{\mathsf{H}} = \boldsymbol{A}_i^{-1} \in \mathbb{C}^{N \times M}$ is the demixing matrix and $\cdot^{\mathsf{H}}$ denotes the Hermitian transpose.

ILRMA assumes that the separated signal y_{ijn} follows a univariate complex Gaussian distribution whose mean and time-variant variance are zero and r_{ijn}, respectively. It is also assumed that the time-variant variance r_{ijn} has low rankness and is modeled by nonnegative matrix factorization (NMF) [11] as $r_{ijn} = \sum_k t_{ikn} v_{kjn}$, where $t_{ikn} \geq 0$ and $v_{kjn} \geq 0$ are the NMF variables, and $k \in \{1, ..., K\}$ is the index of the NMF basis. The cost function of ILRMA $\mathcal{J}_{\text{ILRMA}}$ is given as the following negative log-likelihood of the observed signals:

$$\mathcal{J}_{\text{ILRMA}} = \sum_{i,j,n} \left(\frac{|\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2}{r_{ijn}} + \log r_{ijn} \right) - J \sum_i \log |\det \boldsymbol{W}_i|^2 + \text{const.}, \tag{7}$$

where const. is the term that does not include $\boldsymbol{w}_{in}$, t_{ikn}, or v_{kjn}. The cost function $\mathcal{J}_{\text{ILRMA}}$ is minimized by alternately updating $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn}. In [5], the update rules for the NMF variables t_{ikn} and v_{kjn} are derived on the basis of the MM algorithm. For the update of the demixing matrix $\boldsymbol{W}_i$, iterative projection (IP) [12] is applied. IP also guarantees the monotonic non-increase in the cost function $\mathcal{J}_{\text{ILRMA}}$. Therefore, the total update rules for $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn} also guarantee the monotonic non-increase in the cost function $\mathcal{J}_{\text{ILRMA}}$.

C. SR-ILRMA [9]

To improve the separation performance of ILRMA, SR-ILRMA uses prior information about the source directions and the microphone array geometry as a regularizer in ILRMA. Note that SR-ILRMA is not a fully blind method.

There are two types of regularizer in conventional SR-ILRMA [9], [13]. In [9], the regularizer $\mathcal{R}_1$ is defined as

$$\mathcal{R}_1 = J \sum_{i,n} \mu_{in}^{(\text{SR1})} \|\boldsymbol{w}_{in} - \hat{\boldsymbol{w}}_{in}\|_2^2, \tag{8}$$

where $\mu_{in}^{(\text{SR1})}$ is the weight of the regularizer and $\hat{\boldsymbol{w}}_{in} \in \mathbb{C}^{M}$ is the prior demixing filter corresponding to $\boldsymbol{w}_{in}$. On the other hand, in [13], the regularizer $\mathcal{R}_2$ is defined as

$$\mathcal{R}_2 = J \sum_{i,n} \mu_{in}^{(\text{SR2})} |\boldsymbol{w}_{in}^{\mathsf{H}} \hat{\boldsymbol{a}}_{in^{(t)}} - \delta_{nn^{(t)}}|^2, \tag{9}$$

where $\mu_{in}^{(\text{SR2})}$ is the weight of the regularizer, and $\hat{\boldsymbol{a}}_{in^{(t)}} \in \mathbb{C}^{M}$ is the prior SV of the target source. Here, $n^{(t)}$ is the index of the target source and $\delta_{nn^{(t)}}$ is the Kronecker delta.

For the convenience of the proposed method, unlike (8) and (9), we define the regularizer $\mathcal{R}$ as

$$\mathcal{R} = J \sum_{i,n,n'} \mu_{inn'} |\boldsymbol{w}_{in}^{\mathsf{H}} \hat{\boldsymbol{a}}_{in'} - \delta_{nn'}|^2, \tag{10}$$

where $\mu_{inn'}$ is the weight of the regularizer, $\hat{\boldsymbol{A}}_i = (\hat{\boldsymbol{a}}_{i1}, ..., \hat{\boldsymbol{a}}_{iN}) \in \mathbb{C}^{M \times N}$ is the prior mixing matrix corresponding to $\boldsymbol{A}_i$, and $\hat{\boldsymbol{a}}_{in}$ is the prior SV of the nth source. Note that the regularizer itself is the same as that proposed in [8], [14]. In this case, the cost function $\mathcal{J}_{\text{SR}}$ is defined as

$$\mathcal{J}_{\text{SR}} = \mathcal{J}_{\text{ILRMA}} + \mathcal{R}. \tag{11}$$

In SR-ILRMA, the cost function $\mathcal{J}_{\text{SR}}$ is minimized by alternately updating $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn}. Since the regularizer $\mathcal{R}$ does not include t_{ikn} or v_{kjn}, the update rules for t_{ikn} and v_{kjn} can be derived in the same manner as those in ILRMA. For the update of $\boldsymbol{W}_i$, vectorwise coordinate descent (VCD) [15] can be applied instead of IP because $\mathcal{J}_{\text{SR}}$ consists of a quadratic form, the logarithm of the determinant, and a linear term of

$\boldsymbol{w}_{in}$. Therefore, the update rules for $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn} in SR-ILRMA are as follows:

$$t_{ikn} \leftarrow t_{ikn}\sqrt{\frac{\sum_j v_{kjn}|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2/(\sum_{k'} t_{ik'n}v_{k'jn})^2}{\sum_j(v_{kjn}/\sum_{k'} t_{ik'n}v_{k'jn})}}, \quad (12)$$

$$v_{kjn} \leftarrow v_{kjn}\sqrt{\frac{\sum_i t_{ikn}|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2/(\sum_{k'} t_{ik'n}v_{k'jn})^2}{\sum_i(t_{ikn}/\sum_{k'} t_{ik'n}v_{k'jn})}}, \quad (13)$$

$$\hat{\boldsymbol{D}}_{in} \leftarrow \frac{1}{J}\sum_j \frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{\sum_k t_{ikn}v_{kjn}} + \sum_{n'}\mu_{inn'}\hat{\boldsymbol{a}}_{in'}\hat{\boldsymbol{a}}_{in'}^{\mathsf{H}}, \quad (14)$$

$$\boldsymbol{u}_{in} \leftarrow (\boldsymbol{W}_i\hat{\boldsymbol{D}}_{in})^{-1}\hat{\boldsymbol{a}}_{in}, \quad (15)$$

$$\hat{\boldsymbol{u}}_{in} \leftarrow \mu_{inn}\hat{\boldsymbol{D}}_{in}^{-1}\boldsymbol{u}_{in}, \quad (16)$$

$$h_{in} \leftarrow \boldsymbol{u}_{in}^{\mathsf{H}}\hat{\boldsymbol{D}}_{in}\boldsymbol{u}_{in}, \quad (17)$$

$$\hat{h}_{in} \leftarrow \boldsymbol{u}_{in}^{\mathsf{H}}\hat{\boldsymbol{D}}_{in}\hat{\boldsymbol{u}}_{in}, \quad (18)$$

$$\chi_{in} \leftarrow \begin{cases} \dfrac{1}{\sqrt{h_{in}}}, & (\text{if } \hat{h}_{in}=0) \\ \dfrac{\hat{h}_{in}}{2h_{in}}\left[\sqrt{1+\dfrac{4h_{in}}{|\hat{h}_{in}|^2}}-1\right], & (\text{otherwise}) \end{cases} \quad (19)$$

$$\boldsymbol{w}_{in} \leftarrow \chi\boldsymbol{u}_{in} + \hat{\boldsymbol{u}}_{in}. \quad (20)$$

VCD also guarantees the monotonic non-increase in the cost function $\mathcal{J}_{\text{SR}}$. Therefore, the total update rules for $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn} also guarantee the monotonic non-increase in the cost function $\mathcal{J}_{\text{SR}}$.

III. Proposed method

A. Motivation

SR-ILRMA requires prior information about the source directions and the microphone array geometry to precalculate $\hat{\boldsymbol{w}}_{in}$ or $\hat{\boldsymbol{a}}_{in}$, i.e., SR-ILRMA is not a fully blind method. However, a fully blind method is still preferable because of its broad range of applications. In this paper, we propose a new fully blind method to estimate SV $\hat{\boldsymbol{a}}_{in}$ simultaneously with the demixing matrix and the NMF variables by minimizing the cost function $\mathcal{J}_{\text{SR}}$.

If we directly optimize (11) with respect to $\hat{\boldsymbol{a}}_{in'}$ considering that the current $\boldsymbol{W}_i$, t_{ikn}, and v_{kjn} are fixed, over-fitting probably occurs. Therefore, we need to introduce a small number of other parameters to model the SV. Then, we use TDOAs, which implicitly have information about the source directions and the microphone array geometry, as the parameters for SV. To guarantee the monotonic non-increase in the proposed cost function during the simultaneous estimation of TDOAs and other parameters, we derive the update rules on the basis of the MM algorithm. We also investigate the method of initializing TDOAs for the proposed method.

B. Modeling

Assuming a plane wave arrival model, $\hat{\boldsymbol{a}}_{in}$ can be represented as

$$\hat{\boldsymbol{a}}_{in}(\boldsymbol{\tau}_n) = (e^{-\mathrm{j}\omega_i\tau_{1n}}, \cdots, e^{-\mathrm{j}\omega_i\tau_{Mn}})^{\mathsf{T}}, \quad (21)$$

where $\boldsymbol{\tau}_n = (\tau_{1n}, \cdots, \tau_{Mn})^{\mathsf{T}} \in \mathbb{R}^M$, j is the imaginary unit, and ω_i is the normalized angular frequency corresponding to the frequency bin i. Here, τ_{mn} denotes the TDOA.

Introducing (21) into (11), we define the proposed cost function $\mathcal{J}_{\text{proposed}}$ as

$$\mathcal{J}_{\text{proposed}} = \mathcal{J}_{\text{ILRMA}} + \sum_{i,n,n'}\mu_{inn'}\left|\boldsymbol{w}_{in}^{\mathsf{H}}\hat{\boldsymbol{a}}_{in'}(\boldsymbol{\tau}_{n'}) - \delta_{nn'}\right|^2. \quad (22)$$

We consider optimizing (22) by alternating iterative updates of $\boldsymbol{w}_{in}$, t_{ikn}, v_{kjn}, and $\boldsymbol{\tau}_{n'}$. Since $\boldsymbol{\tau}_{n'}$ is fixed during updating $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn}, the update rules for $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn} are the same as those in SR-ILRMA. Therefore, we consider the update rules for $\boldsymbol{\tau}_{n'}$ hereafter.

C. Derivation of update rules for proposed method

$\boldsymbol{\tau}_{n'}$ is only involved in the regularizer of (22); thus, we consider minimizing the following cost function $\mathcal{J}$:

$$\mathcal{J} = \sum_{i,n,n'}\mu_{inn'}\left|\boldsymbol{w}_{in}^{\mathsf{H}}\hat{\boldsymbol{a}}_{in'}(\boldsymbol{\tau}_{n'}) - \delta_{nn'}\right|^2. \quad (23)$$

We substitute (21) into (23) and rearrange the formula as

$$\begin{aligned}\mathcal{J} = \sum_{i,n,n'}\mu_{inn'}\Bigg\{&\sum_{m,m'} w_{inm}^* e^{-\mathrm{j}\omega_i\tau_{mn'}} w_{inm'} e^{\mathrm{j}\omega_i\tau_{m'n'}} \\ &- \delta_{nn'}\sum_m (w_{inm}^* e^{-\mathrm{j}\omega_i\tau_{mn'}} \\ &+ w_{inm'}e^{\mathrm{j}\omega_i\tau_{m'n'}}) + \delta_{nn'}\Bigg\}, \quad (24)\end{aligned}$$

$$\begin{aligned}&= 2\sum_{i,n,n'}\mu_{inn'}\left(\sum_{m<m'}|w_{inm}||w_{inm'}|\cos\theta_{inn'mm'}\right) \\ &\quad + 2\sum_{i,n'}\mu_{in'n'}\left(\sum_m |w_{in'm}|\cos\phi_{in'm}\right) + \text{const.}, \quad (25)\end{aligned}$$

where $\cdot^*$ denotes the complex conjugation and const. is the term that does not include $\boldsymbol{\tau}_{n'}$. Here, $\theta_{inn'mm'}$ and $\phi_{in'm}$ are respectively defined as

$$\theta_{inn'mm'} = \omega_i\tau_{mn'} - \omega_i\tau_{m'n'} + \angle w_{inm} - \angle w_{inm'}, \quad (26)$$

$$\phi_{in'm} = \omega_i\tau_{mn'} + \angle w_{in'm} + \pi, \quad (27)$$

where $\angle\cdot$ denotes the argument of a complex number.

Since (25) is the sum of cosine functions and the coefficients of the cosine functions are always positive, we can use an auxiliary function for the cosine function to design the auxiliary function for the cost function $\mathcal{J}$. For any ϑ and $\tilde{\vartheta}$, the following inequality [16] is satisfied:

$$\begin{aligned}\cos\vartheta \le \frac{1}{2}\mathrm{sinc}\big(g(\tilde{\vartheta})\big)\big(\vartheta - \tilde{\vartheta} + g(\tilde{\vartheta})\big)^2 \\ + \cos\tilde{\vartheta} + \frac{1}{2}g(\tilde{\vartheta})\sin\tilde{\vartheta}, \quad (28)\end{aligned}$$

where $\mathrm{sinc}(\varphi) = \sin(\varphi)/\varphi$ denotes the unnormalized sinc function and $g(\cdot) = \mathrm{mod}_{2\pi}(\cdot) - \pi$. Here, $\mathrm{mod}_{2\pi}(\cdot)$ denotes the modulo operation with 2π. The equality in (28) holds when

$\vartheta = \tilde{\vartheta}$, and the right-hand side of (28) serves as an auxiliary function for $\cos\vartheta$ with the auxiliary variable $\tilde{\vartheta}$.

Using (28), we can design the auxiliary function $\mathcal{J}^+$ for $\mathcal{J}$ as

$$\begin{aligned}\mathcal{J}^+ = &\sum_{i,n,n'} \mu_{inn'} \left\{ \sum_{m<m'} \alpha_{inn'mm'} \Big(\omega_i \Delta\tau_{mn'} \right. \\ &\left. - \omega_i \Delta\tau_{m'n'} + g(\tilde{\theta}_{inn'mm'}) \Big)^2 \right\} \\ &+ \sum_{i,n'} \mu_{in'n'} \left\{ \sum_m \beta_{in'm} \Big(\omega_i \Delta\tau_{mn'} + g(\tilde{\phi}_{in'm}) \Big)^2 \right\} \\ &+ \text{const.}, \end{aligned} \tag{29}$$

where const. is the term that does not include $\boldsymbol{\tau}_{n'}$, and $\alpha_{inn'mm'}$, $\beta_{in'm}$, and $\Delta\tau_{mn'}$ are respectively defined as

$$\alpha_{inn'mm'} = |w_{inm}||w_{inm'}|\mathrm{sinc}(g(\tilde{\theta}_{inn'mm'})), \tag{30}$$

$$\beta_{in'm} = |w_{in'm}|\mathrm{sinc}(g(\tilde{\phi}_{in'm})), \tag{31}$$

$$\Delta\tau_{mn'} = \tau_{mn'} - \tilde{\tau}_{mn'}, \tag{32}$$

where $\{\tilde{\tau}_{mn'}\}$ is the set of auxiliary variables, and $\tilde{\theta}_{inn'mm'}$ and $\tilde{\phi}_{in'm}$ are defined by substituting $\tau_{mn'} = \tilde{\tau}_{mn'}$ into (26) and (27). The equality in (29) holds when $\tau_{mn'} = \tilde{\tau}_{mn'}$ for all m and n'.

Since $\mathcal{J}^+$ is a quadratic function of $\Delta\tau_{mn'}$, we can deform $\mathcal{J}^+$ by using a quadratic form and linear term of $\Delta\boldsymbol{\tau}_{n'}$ as

$$\mathcal{J}^+ = \sum_{n'} \left(\Delta\boldsymbol{\tau}_{n'}^{\mathsf{T}} \boldsymbol{Q}_{n'} \Delta\boldsymbol{\tau}_{n'} - 2\Delta\boldsymbol{\tau}_{n'}^{\mathsf{T}} \boldsymbol{b}_{n'} \right) + \text{const.}, \tag{33}$$

where const. is the term that does not include $\boldsymbol{\tau}_{n'}$. Here, $\Delta\boldsymbol{\tau}_{n'}$, $\boldsymbol{Q}_{n'} \in \mathbb{R}^{M\times M}$, and $\boldsymbol{b}_{n'} \in \mathbb{R}^M$ are defined as

$$\tilde{\boldsymbol{\tau}}_{n'} = (\tilde{\tau}_{1n'}, \cdots, \tilde{\tau}_{Mn'})^{\mathsf{T}} \in \mathbb{R}^M, \tag{34}$$

$$\Delta\boldsymbol{\tau}_{n'} = \boldsymbol{\tau}_{n'} - \tilde{\boldsymbol{\tau}}_{n'}, \tag{35}$$

$$\begin{aligned}(\boldsymbol{Q}_{n'})_{mm'} &= \begin{cases} \displaystyle\sum_{i,n} \mu_{inn'} \omega_i^2 \left(\sum_{m''\neq m} \alpha_{inn'mm''} \right) & (m = m') \\ \qquad + \displaystyle\sum_i \mu_{in'n'} \omega_i^2 \beta_{in'm}, & \\ \displaystyle - \sum_{i,n} \mu_{inn'} \omega_i^2 \alpha_{inn'mm'}, & (m \neq m') \end{cases} \\ &= \begin{cases} \displaystyle\sum_{i,n} \mu_{inn'} \omega_i^2 \left(\sum_{m''\neq m} \alpha_{inn'mm''} \right) & (m = m') \\ \qquad + \displaystyle\sum_i \mu_{in'n'} \omega_i^2 \beta_{in'm}, & \\ \displaystyle - \sum_{i,n} \mu_{inn'} \omega_i^2 \alpha_{inn'mm'}, & (m \neq m') \end{cases} \end{aligned} \tag{36}$$

$$\begin{aligned}(\boldsymbol{b}_{n'})_m = &- \sum_{i,n} \mu_{inn'} \omega_i \left(\sum_{m'} \alpha_{inn'mm'} g(\tilde{\theta}_{inn'mm'}) \right) \\ &- \sum_i \mu_{in'n'} \omega_i \beta_{in'm} g(\tilde{\phi}_{in'm}), \end{aligned} \tag{37}$$

where $(\cdot)_{mm'}$ denotes the (m,m')th element of a matrix.

Since (33) is separable for each n', the auxiliary function $\mathcal{J}^+$ can be minimized by solving $\partial\mathcal{J}^+/\partial\boldsymbol{\tau}_{n'} = \boldsymbol{0}_M$ for all n'. Here, $\boldsymbol{0}_M \in \mathbb{R}^M$ denotes the M-dimensional zero vector. Therefore, the MM-algorithm-based update rules for $\boldsymbol{\tau}_n'$ are derived as

$$\tilde{\boldsymbol{\tau}}_{n'} \leftarrow \boldsymbol{\tau}_{n'}, \tag{38}$$

$$\boldsymbol{\tau}_{n'} \leftarrow \tilde{\boldsymbol{\tau}}_{n'} + \boldsymbol{Q}_{n'}^{\dagger} \boldsymbol{b}_{n'}, \tag{39}$$

where $\cdot^{\dagger}$ denotes the Moore–Penrose inverse. Note that we use the Moore–Penrose inverse considering if $\boldsymbol{Q}_{n'}$ may not be regular.

Since we can use the same update rules as (12)–(20) for $\boldsymbol{w}_{in}$, t_{ikn}, and v_{kjn}, the total update rules for $\boldsymbol{w}_{in}$, t_{ikn}, v_{kjn}, and $\boldsymbol{\tau}_n'$ in the proposed method are

$$t_{ikn} \leftarrow t_{ikn} \sqrt{\frac{\sum_j v_{kjn} |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2 / (\sum_{k'} t_{ik'n} v_{k'jn})^2}{\sum_j (v_{kjn} / \sum_{k'} t_{ik'n} v_{k'jn})}}, \tag{40}$$

$$v_{kjn} \leftarrow v_{kjn} \sqrt{\frac{\sum_i t_{ikn} |\boldsymbol{w}_{in}^{\mathsf{H}} \boldsymbol{x}_{ij}|^2 / (\sum_{k'} t_{ik'n} v_{k'jn})^2}{\sum_i (t_{ikn} / \sum_{k'} t_{ik'n} v_{k'jn})}}, \tag{41}$$

$$\hat{\boldsymbol{D}}_{in} \leftarrow \frac{1}{J} \sum_j \frac{\boldsymbol{x}_{ij} \boldsymbol{x}_{ij}^{\mathsf{H}}}{\sum_k t_{ikn} v_{kjn}} + \sum_{n'} \mu_{inn'} \hat{\boldsymbol{a}}_{in'} \hat{\boldsymbol{a}}_{in'}^{\mathsf{H}}, \tag{42}$$

$$\boldsymbol{u}_{in} \leftarrow (\boldsymbol{W}_i \hat{\boldsymbol{D}}_{in})^{-1} \hat{\boldsymbol{a}}_{in}, \tag{43}$$

$$\hat{\boldsymbol{u}}_{in} \leftarrow \mu_{inn} \hat{\boldsymbol{D}}_{in}^{-1} \boldsymbol{u}_{in}, \tag{44}$$

$$h_{in} \leftarrow \boldsymbol{u}_{in}^{\mathsf{H}} \hat{\boldsymbol{D}}_{in} \boldsymbol{u}_{in}, \tag{45}$$

$$\hat{h}_{in} \leftarrow \boldsymbol{u}_{in}^{\mathsf{H}} \hat{\boldsymbol{D}}_{in} \hat{\boldsymbol{u}}_{in}, \tag{46}$$

$$\chi_{in} \leftarrow \begin{cases} \dfrac{1}{\sqrt{h_{in}}}, & (\text{if } \hat{h}_{in} = 0) \\ \dfrac{\hat{h}_{in}}{2h_{in}} \left[\sqrt{1 + \dfrac{4h_{in}}{|\hat{h}_{in}|^2}} - 1 \right] & (\text{otherwise}) \end{cases} \tag{47}$$

$$\boldsymbol{w}_{in} \leftarrow \chi \boldsymbol{u}_{in} + \hat{\boldsymbol{u}}_{in} \tag{48}$$

$$\tilde{\boldsymbol{\tau}}_{n'} \leftarrow \boldsymbol{\tau}_{n'}, \tag{49}$$

$$\boldsymbol{\tau}_{n'} \leftarrow \tilde{\boldsymbol{\tau}}_{n'} + \boldsymbol{Q}_{n'}^{\dagger} \boldsymbol{b}_{n'}, \tag{50}$$

$$\hat{\boldsymbol{a}}_{in'} \leftarrow \hat{\boldsymbol{a}}_{in'}(\boldsymbol{\tau}_{n'}), \tag{51}$$

where $\boldsymbol{Q}_{n'}$ and $\boldsymbol{b}_{n'}$ are calculated by substituting (49) into (36) and (37), respectively. The update rules for $\boldsymbol{w}_{in}$, t_{ikn}, v_{kjn}, and $\boldsymbol{\tau}_n'$ (40)–(51) guarantee the monotonic non-increase in the cost function $\mathcal{J}_{\text{proposed}}$.

D. Initialization method based on impulse responses

The cost function $\mathcal{J}$ is the sum of cosine functions, and the cosine function has an infinite number of local optimal solutions. Thus, the proposed method seems to be highly affected by the initial values of τ_{mn}. Then, we consider the initialization method for τ_{mn}. Instead of $\mathcal{J}$, we consider $\sum_i \|\hat{\boldsymbol{A}}_i - \boldsymbol{W}_i^{-1}\|_{\mathrm{F}}^2$ for the initialization, where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm. On the basis of Parseval's theorem, the following equation holds:

$$\begin{aligned}&\sum_i \|\hat{\boldsymbol{A}}_i - \boldsymbol{W}_i^{-1}\|_{\mathrm{F}}^2 \\ &= \gamma \sum_{m,n} \Big\| \mathrm{IDFT}\big[\{e^{-\mathrm{j}\omega_i \tau_{mn}}\}\big] - \mathrm{IDFT}\big[\{(\boldsymbol{W}_i^{-1})_{mn}\}\big] \Big\|_2^2, \end{aligned} \tag{52}$$

where γ is the length of the STFT window and $\mathrm{IDFT}[\{z_i\}] \in \mathbb{R}^{\gamma}$ denotes the inverse discrete Fourier transform of the complex spetrum whose ith frequency bin is z_i

When τ_{mn} is discrete, the $(\tau_{mn}+1)$th element of $\mathrm{IDFT}[\{e^{-\mathrm{j}\omega_i\tau_{mn}}\}]$ becomes one and the other elements become zero. Therefore, to minimize (52), we choose the τ_{mn}

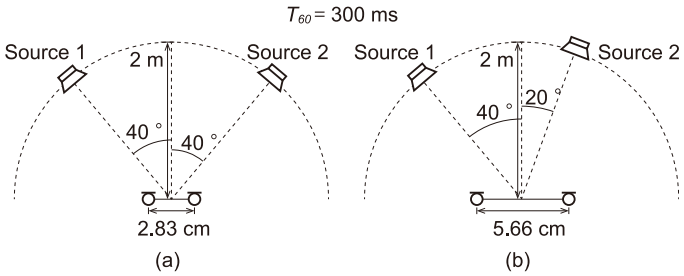

Fig. 1. Recording conditions of impulse responses.

corresponding to the index where $\mathrm{IDFT}\big[\{(\boldsymbol{W}_i^{-1})_{mn}\}\big]$, which means the impulse response (IR) calculated from $\boldsymbol{W}_i$, takes the maximum value. Then, we use this optimal τ_{mn} of the pseudo-problem as the initial value of the proposed cost function.

IV. Experiments

First, we considered a subproblem to estimate TDOAs in the situation where the demixing matrix was fixed and conducted a preliminary experiment to investigate the initial value dependence of the proposed method. Next, we evaluated the separation performance of the proposed method through a numerical experiment. Both experiments were conducted using two sources and two microphones.

A. Experimental conditions

We used three pairs of instruments (bass/drums, drums/vocal, vocal/bass) for each of the six tracks in the DSD100 dataset [17] as dry sources (total 18 pairs). Then, each dry source was convolved with two conditions of impulse responses in the RWCP database [18] (see Fig. 1). The convolved signals were mixed so that the input signal-to-noise ratio became $0\,\mathrm{dB}$. The total number of the observed signals was $18 \times 2 = 36$. The sampling rate was $16\,\mathrm{kHz}$. STFT was performed using a 256-ms-long Hamming window with a shift length of $64\,\mathrm{ms}$.

B. Initial value dependence

To investigate the initial value dependence of the proposed method, we considered the subproblem to estimate the TDOAs in the situation where the demixing matrix $\boldsymbol{W}_i$ was fixed. First, we obtained the demixing matrix by ILRMA as the preprocess. Then, the demixing matrix was fixed, and only the updates related to TDOAs were iteratively performed to observe the behavior of the cost function $\mathcal{J}$ during iteration.

In ILRMA as the preprocess, the demixing matrix was initialized with the identity matrix, and the NMF variables t_{ikn} and v_{kjn} were initialized with random numbers from a uniform distribution on $[0, 1]$. We used 10 random seeds for each observed signal, resulting in 360 demixing matrices. The number of iterations was set to 100. In the estimation of TDOAs, we compared two initialization methods: (i) random sampling and (ii) proposed initialization described in Section III-D (IR-based initialization method). In random sampling, 10 different initial values were uniformly sampled from $[-5, 5]$. The weighting coefficient $\mu_{inn'}$ for the regularizer was set as $\mu_{inn'} = 1$ $(i \neq 1, I)$ and $\mu_{inn'} = 0$ $(i = 1, I)$. The number of iterations was set to 200.

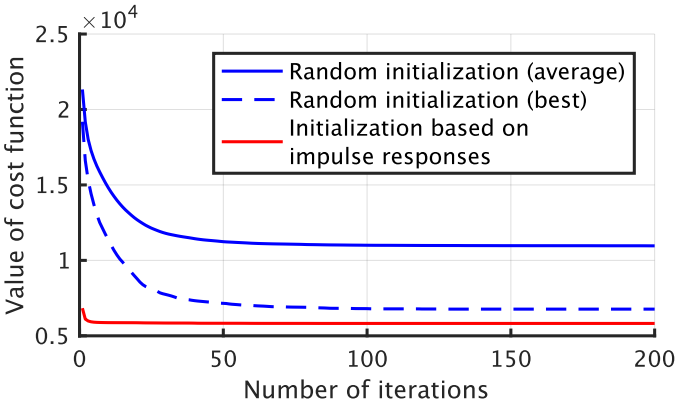

Fig. 2. Behavior of the cost function (23) for each initialization method. The average behavior for random initial values is shown by the solid blue line, the behavior for the random initial values that resulted in the smallest final value of cost function is shown by the dashed blue line, and the behavior for the initial value calculated by the IR-based initialization method is shown by the solid red line.

Fig. 2 shows the behavior of the cost function for the following three cases: the solid blue line shows the average behavior for the 10 random initial values, the dashed blue line shows the behavior for the initial value that resulted in the smallest final value of cost function among the 10 random initial values, and the solid red line shows the behavior for the initial value calculated by the IR-based initialization method.

It can be confirmed that the IR-based initialization method achieves the smallest value of the cost function even at the end and a faster convergence. This suggests that the IR-based initialization method yields the initial value close to the global optimum of the subproblem.

C. Comparison of separation performance

To confirm the efficiency of the proposed method, we compared the following three methods.

- **ILRMA [5]**
- **Proposed method (random init.)**: TDOAs are initialized by a random sampling.
- **Proposed method (IR-based init.)**: TDOAs are initialized by the IR-based initialization method using the demixing matrix precomputed by ILRMA.

In ILRMA, initialization was the same as that described in Section IV-B. The number of iterations was set to 200. In the proposed method (random init.), the demixing matrix and NMF variables were initialized in the same manner as in ILRMA in Section IV-B. TDOAs were initialized with random values from the uniform distribution on $[-5, 5]$. The number of iterations was also set to 200 as in baseline ILRMA. In the proposed method (IR-based init.), to start with better initial values of TDOAs, we first performed 20 iterations of ILRMA to obtain the demixing matrix as the preprocess. TDOAs were then initialized using the IR-based initialization method, followed by 180 iterations of the proposed method.

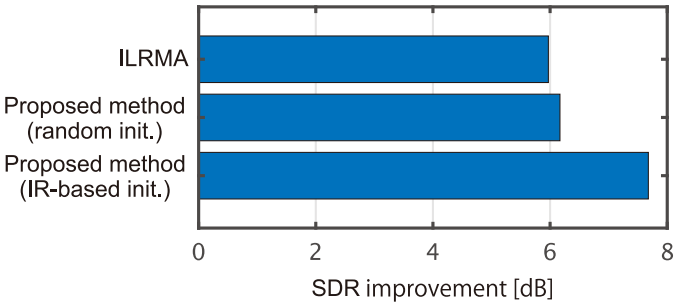

Fig. 3. Average SDR improvement for each method.

The preprocessing ILRMA was initialized in the same manner as in Section IV-B. In the subsequent proposed method, the NMF variables were again initialized as described in Section IV-B. The demixing matrix was initialized with the inverse of the mixing matrix $\hat{\boldsymbol{A}}_i$ computed from the TDOAs obtained by the IR-based initialization method.

Since the plane wave arrival model does not account for reverberation, strong regularization may affect the final separation performance. To mitigate this issue, the weighting coefficient $\mu_{inn'}^{(l)}$ for the regularizer in the lth update was decayed as

$$\mu_{inn'}^{(l)} = \begin{cases} 0, & (i = 1, I) \\ \mu_o \max\left[1/2 - l/L, 0\right], & (i \neq 1, I) \end{cases} \tag{53}$$

where L is the number of iterations for proposed methods, i.e., $L = 200$ in proposed method (random init.) and $L = 180$ in proposed method (IR-based init.). The value of μ_o was determined by a preliminary experiment and set to 10^{-3}.

In each method, we used 10 random seeds for each of the observed signals (total 360 trials) and calculated the average of source-to-distortion ratio (SDR) [19] improvement as the evaluation metric for the separation performance.

Fig. 3 shows the average SDR improvement for each method. It can be confirmed that the proposed method achieves better performance than ILRMA. Additionally, the proposed method (IR-based init.) achieves the largest SDR improvement. Therefore, we can confirm the effectiveness of the proposed fully blind method.

V. Conclusion

In this paper, we proposed a new fully blind method that models SVs using TDOAs and simultaneously estimates TDOAs, the demixing matrix, and NMF variables. To guarantee the monotonic non-increase in the cost function, we derived the MM-algorithm-based update rules. We also investigated the method of initializing TDOAs for the proposed cost function. Numerical experiments confirmed that the proposed method achieves higher separation performance than ILRMA. Additionally, we confirmed the effectiveness of the IR-based initialization method for the proposed method.

References

[1] H. Sawada, N. Ono, H. Kameoka, D. Kitamura, and H. Saruwatari, "A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF," *APSIPA Trans. SIP*, vol. 8, no. e12, pp. 1–14, 2019.

[2] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1, pp. 21–34, 1998.

[3] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Proc. ICA*, 2006, pp. 601–608.

[4] T. Kim, H. T. Attias, S. Y. Lee, and T. W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Trans. ASLP*, vol. 15, no. 1, pp. 70–79, 2007.

[5] D. Kitamura, N. Ono, H. Sawada, H. Kameoka, and H. Saruwatari, "Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization," *IEEE/ACM Trans. ASLP*, vol. 24, pp. 1626–1641, 2016.

[6] D. Kitamura, N. Ono, and H. Saruwatari, "Experimental analysis of optimal window length for independent low-rank matrix analysis," in *Proc. EUSIPCO*, 2017, pp. 1170–1174.

[7] H. Saruwatari, T. Kawamura, T. Nishikawa, A. Lee, and K. Shikano, "Blind source separation based on a fast-convergence algorithm combining ICA and beamforming," *IEEE Trans. ASLP*, vol. 14, no. 2, pp. 666–678, 2006.

[8] L. Li and K. Koishida, "Geometrically constrained independent vector analysis for directional speech enhancement," in *Proc. ICASSP*, 2020, pp. 846–850.

[9] Y. Mitsui, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, "Vectorwise coordinate descent algorithm for spatially regularized independent low-rank matrix analysis," in *Proc. ICASSP*, 2018, pp. 746–750.

[10] D. R. Hunter and K. Lange, "Quantile regression via an MM algorithm," *JCGS*, vol. 9, no. 1, pp. 60–77, 2000.

[11] D. D. Lee and H. S. Seung, "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, no. 6755, pp. 788–791, 1999.

[12] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. WASPAA*, 2011, pp. 189–192.

[13] Y. Ishikawa, K. Konaka, T. Nakamura, N. Takamune, and H. Saruwatari, "Real-time speech extraction using spatially regularized independent low-rank matrix analysis and rank-constrained spatial covariance matrix estimation," in *Proc. HSCMA*, 2024.

[14] L. C. Parra and C. V. Alvino, "Geometric source separation: Merging convolutive source separation with geometric beamforming," *IEEE TSAP*, vol. 10, no. 6, pp. 352–362, 2002.

[15] N. Makishima, Y. Mitsui, N. Takamune, *et al.*, "Independent deeply learned matrix analysis with automatic selection of stable microphone-wise update and fast sourcewise update of demixing matrix," *Signal Processing*, vol. 178, p. 107 753, 2021.

[16] K. Yamaoka, R. Scheibler, N. Ono, and Y. Wakabayashi, "Sub-sample time delay estimation via auxiliary-function-based iterative updates," in *Proc. WASPAA*, 2019, pp. 130–134.

[17] A. Liutkus, F. R. Stöter, Z. Rafii, *et al.*, "The 2016 signal separation evaluation campaign," in *Proc. LVA/ICA*, 2017, pp. 323–332.

[18] S. Nakamura, K. Hiyane, F. Asano, T. Nishiura, and T. Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," in *Proc. LREC*, 2000.

[19] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. ASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.