Theoretical Formulation of Online Independent Vector Analysis Using Framewise Probabilistic Generative Model

Yuto Ishikawa, Norihiro Takamune, Kouei Yamaoka, Hiroshi Saruwatari
Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
yuto_ishikawa.jp@ieee.org, {norihiro_takamune, kouei_yamaoka, hiroshi_saruwatari}@ipc.i.u-tokyo.ac.jp

***Abstract*—Real-time blind source separation is an important technique for many real-world applications. One of the representative methods is an online extension of independent vector analysis (IVA), which is called online IVA (OIVA). However, the optimization problem of OIVA has not been theoretically discussed thus far. In this paper, we introduce a framewise probabilistic generative model to formulate OIVA on the basis of statistical independence and derive the update rules. The theoretical background of the conventional OIVA is also derived as the limiting case of the proposed OIVA. Simulation experiments demonstrate that the proposed OIVA achieves faster convergence in separation performance than the conventional OIVA.**

***Index Terms*—independent vector analysis, online independent vector analysis, online blind source separation**

I. Introduction

Multichannel blind source separation (BSS) is a technique to separate each source signal from mixtures recorded by a microphone array without any prior information [1]. Under (over-)determined conditions, where the number of microphones is greater than or equal to that of sources, a major approach is to estimate the demixing matrix so that the separated signals become statistically independent of each other. One of the commonly used methods based on this approach is independent vector analysis (IVA) [2], [3]. By introducing a model that considers higher-order dependencies between frequency components, IVA can achieve high source separation performance. For fast and numerically stable estimation, auxiliary-function-based IVA (AuxIVA) has been proposed [4]. AuxIVA is based on the auxiliary function method [5] and can be performed without hyperparameter tuning unlike in [2], [3].

In real-world applications such as hearing aids and robot dialogue systems, real-time BSS methods are necessary for smooth communication. As one of such methods, an online extension of AuxIVA, which is called online AuxIVA (OIVA), has been proposed [6], [7]. In OIVA, the update algorithm of AuxIVA is extended to process the observed signal online, i.e., in a frame-by-frame manner. It has been experimentally shown that OIVA performs well in online scenarios. However, in the conventional OIVAs, the update rule of AuxIVA is heuristically extended for online implementation, and a corresponding optimization problem has not been revealed.

In this paper, we introduce a framewise probabilistic generative model to formulate OIVA as an optimization problem based on statistical independence. The proposed framewise formulation can also represent offline BSS problems by adjusting a hyperparameter, making it possible to represent both offline and online BSS problems. We also show that the update rules of the proposed OIVAs become identical to those of the conventional OIVAs in a limiting case where an infinitely long time has passed. Moreover, our proposed method executes some suitable estimation in situations where the number of available frames is small, such as at the start of observation. We conducted two experiments: (i) all sources were stationary and (ii) a source moved. Through these experiments, we confirmed that the proposed OIVAs achieve faster convergence in source separation performance than the conventional OIVAs.

II. Related works

A. IVA

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 $n \in \{1, ..., N\}$ denote the indices of the frequency bins, time frames, microphones, and sources, respectively, and $^\mathsf{T}$ represents the transpose. With the assumptions that each source is a point source and the room reverberation is sufficiently shorter than the window size of an STFT, the following instantaneous mixing in the time-frequency domain approximately holds:

$$\boldsymbol{x}_{ij} = \boldsymbol{A}_i \boldsymbol{s}_{ij}, \tag{1}$$

where $\boldsymbol{A}_i = (\boldsymbol{a}_{i1}, ..., \boldsymbol{a}_{iN}) \in \mathbb{C}^{M \times N}$ is the mixing matrix, which represents the time-invariant spatial characteristics of the mixing system, and $\boldsymbol{a}_{in}$ is the steering vector of the nth source. If $M = N$ and $\boldsymbol{A}_i$ is regular, there exists the inverse of the mixing matrix, $\boldsymbol{W}_i = (\boldsymbol{w}_{i1}, ..., \boldsymbol{w}_{iN})^\mathsf{H} = \boldsymbol{A}_i^{-1}$, where $^\mathsf{H}$ represents the Hermitian transpose and $\boldsymbol{W}_i$ is called the demixing matrix. By using $\boldsymbol{W}_i$, we can obtain the separated signals as

$$\boldsymbol{y}_{ij} = \boldsymbol{W}_i \boldsymbol{x}_{ij}. \tag{2}$$

This work was supported by JST Moonshot R&D Grant Number JPMJMS2011 (for algorithm development), YAMAHA Corporation (for implementation), and Tateisi Science and Technology Foundation (for experiments).

On the basis of statistical independence, the objective variable $\boldsymbol{W}_i$ is estimated by minimizing the following negative log-likelihood:

$$\begin{aligned}&-\log p(\{x_{ijm}\}_{i,j,m}^{I,J,M})\\&=-\sum_n \log p(\{y_{ijn}\}_{i,j}^{I,J}) - J\sum_i \log|\det \boldsymbol{W}_i|^2,\end{aligned} \tag{3}$$

where $\{\cdot\}_{i,j,m}^{I,J,M}$ denotes $\{\{\{\cdot\}_{i=1}^{I}\}_{j=1}^{J}\}_{m=1}^{M}$ for simplicity.

In IVA, the separated signals are assumed to have a higher-order correlation between frequency components and to be independently and identically distributed across time frames. In this paper, we assume that the separated signals follow the following spherical multivariate Laplace distribution with a mean of $\boldsymbol{0}_I$ and a scale matrix of $\boldsymbol{E}_I$ as

$$p(\{y_{ijn}\}_i^I; \boldsymbol{0}_I, \boldsymbol{E}_I) \propto \exp\left(-\sqrt{\sum_i |y_{ijn}|^2}\right), \tag{4}$$

where $\boldsymbol{0}_I \in \mathbb{C}^I$ is the zero vector and $\boldsymbol{E}_I \in \mathbb{C}^{I\times I}$ is the identity matrix. By substituting (4) and $y_{ijn} = \boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}$ into (3), we can obtain the cost function of IVA as

$$\mathcal{L}_{\text{IVA}} = \sum_n \sum_{j=1}^{J} \sqrt{\sum_i |\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2} - J\sum_i \log|\det \boldsymbol{W}_i|^2. \tag{5}$$

It is difficult to directly minimize $\mathcal{L}_{\text{IVA}}$ with respect to $\boldsymbol{W}_i$. In AuxIVA [4], an iterative update rule is derived on the basis of the auxiliary function method [5] as follows. By utilizing the relationship between a concave function and its tangent line, $\sqrt{x} \le x/2\sqrt{c} + \sqrt{c}/2$ $(\forall x, c \in \mathbb{R}_{\ge 0})$, we consider the following auxiliary function of $\mathcal{L}_{\text{IVA}}$:

$$\begin{aligned}\tilde{\mathcal{L}}_{\text{IVA}} =& \sum_n \sum_{j=1}^{J}\left(\frac{\sum_i |\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{2r_{jn}} + \frac{r_{jn}}{2}\right)\\&- J\sum_i \log|\det \boldsymbol{W}_i|^2\\=& J\left(\sum_{i,n} \boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{D}_{in}\boldsymbol{w}_{in} - \sum_i \log|\det \boldsymbol{W}_i|^2\right) + \text{const.},\end{aligned} \tag{6}$$

where const. denotes the term independent of $\boldsymbol{w}_{in}$, r_{jn} is a nonnegative auxiliary variable, and $\boldsymbol{D}_{in}$ is a weighted covariance matrix of the observed signal and defined as

$$\boldsymbol{D}_{in} = \frac{1}{J}\sum_{j=1}^{J} \frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{2r_{jn}}. \tag{7}$$

Here, $\tilde{\mathcal{L}}_{\text{IVA}} = \mathcal{L}_{\text{IVA}}$ holds if and only if

$$r_{jn} = \sqrt{\sum_i |\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2} \tag{8}$$

holds. We can minimize $\mathcal{L}_{\text{IVA}}$ by iteratively repeating the following two processes:

- Minimizing $\tilde{\mathcal{L}}_{\text{IVA}}$ with respect to $\boldsymbol{W}_i$
- Updating r_{jn} and $\boldsymbol{D}_{in}$ using (8) and (7), respectively

For minimizing the auxiliary function $\tilde{\mathcal{L}}_{\text{IVA}}$ with respect to the demixing matrix $\boldsymbol{W}_i$, two methods have been proposed and are commonly used: iterative projection (IP) [4] and iterative source steering (ISS) [8]. The update rule using IP is expressed as

$$\boldsymbol{u}_{in} \leftarrow (\boldsymbol{W}_i\boldsymbol{D}_{in})^{-1}\boldsymbol{e}_n, \tag{9}$$

$$\boldsymbol{w}_{in} \leftarrow \boldsymbol{u}_{in}/\sqrt{\boldsymbol{u}_{in}^{\mathsf{H}}\boldsymbol{D}_{in}\boldsymbol{u}_{in}}, \tag{10}$$

where $\boldsymbol{e}_n$ denotes the nth column vector of $\boldsymbol{E}_N$. On the other hand, the update rule using ISS is expressed as

$$v_{inn'} \leftarrow \begin{cases}\frac{\boldsymbol{w}_{in'}^{\mathsf{H}}\boldsymbol{D}_{in'}\boldsymbol{w}_{in}}{\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{D}_{in'}\boldsymbol{w}_{in}}, & (\text{if } n' \neq n)\\ 1 - (\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{D}_{in}\boldsymbol{w}_{in})^{-\frac{1}{2}}, & (\text{if } n' = n)\end{cases} \tag{11}$$

$$\boldsymbol{W}_i \leftarrow \boldsymbol{W}_i - (v_{in1}, ..., v_{inN})^{\mathsf{T}}\boldsymbol{w}_{in}^{\mathsf{H}}. \tag{12}$$

Both update rules are sequentially executed for $n = 1, ..., N$. It is guaranteed that both the update rules based on IP and ISS monotonically nonincrease the cost function $\mathcal{L}_{\text{IVA}}$. Note that in [4], the probabilistic generative model of the separated signal is generalized for a super-Gaussian distribution that includes the spherical multivariate Laplace distribution, and we can generalize the subsequent discussion in a similar way.

In IVA, the scale of y_{ijn} can vary across the frequency bins. To fix the scales of y_{ijn} among all the frequency bins, the projection back method [9] is applied to y_{ijn} after estimating $\boldsymbol{W}_i$.

B. Online AuxIVA

When we simply apply offline AuxIVA, the auxiliary variables are updated each time the demixing matrix is updated; thus, the weighted covariance matrix of the observed signal also needs to be recalculated. As a result, the computational cost for updating the weighted covariance matrix of the observed signal increases with the length of the observed signal. In [6], [7], to reduce the computational cost, when the kth frame of the observed signal is obtained, the approximate weighted covariance matrix of the observed signal $\hat{\boldsymbol{D}}_{in}^{(k)}$ is updated in an autoregressive manner as

$$\hat{\boldsymbol{D}}_{in}^{(k)} \leftarrow \alpha\hat{\boldsymbol{D}}_{in}^{(k-1)} + (1-\alpha)\frac{\boldsymbol{x}_{ik}\boldsymbol{x}_{ik}^{\mathsf{H}}}{2r_{kn}}, \tag{13}$$

where $\alpha \in [0, 1)$ denotes the forgetting factor, and the upper right script $^{(k)}$ indicates the parameter estimated in the kth frame. Here, $\hat{\boldsymbol{D}}_{in}^{(0)}$ is initialized as $\varepsilon\boldsymbol{E}_N$ with the stability parameter ε, which is sufficiently small. In (13), when we update $\hat{\boldsymbol{D}}_{in}^{(k)}$, the auxiliary variables r_{jn} $(j = 1, ..., k-1)$ are fixed and only r_{kn} is updated. By replacing the updates of $\boldsymbol{D}_{in}$ with $\hat{\boldsymbol{D}}_{in}^{(k)}$ in (9)–(12), we can derive both the update rules of OIVA using IP and ISS. In addition, since spatial characteristics are not expected to change abruptly, we utilize the estimate of the demixing matrix from the $(k-1)$th frame as the initial value for the demixing matrix in the kth frame, reducing the number of iterations for $\hat{\boldsymbol{D}}_{in}^{(k)}$ and $\boldsymbol{W}_i$.

Furthermore, in OIVA using IP, the following fast algorithm based on the Sherman–Morrison formula is proposed [6]:

$$\boldsymbol{\eta}_{in} \leftarrow \hat{\boldsymbol{U}}_{in}^{(k-1)} \boldsymbol{x}_{ik}, \tag{14}$$

$$\hat{\boldsymbol{U}}_{in}^{(k)} \leftarrow \frac{1}{\alpha}\left(\hat{\boldsymbol{U}}_{in}^{(k-1)} - \frac{\boldsymbol{\eta}_{in}\boldsymbol{\eta}_{in}^{\mathsf{H}}}{\boldsymbol{x}_{ik}^{\mathsf{H}}\boldsymbol{\eta}_{in} + 2r_{kn}\alpha/(1-\alpha)}\right), \tag{15}$$

$$\boldsymbol{u}_{in} \leftarrow \hat{\boldsymbol{U}}_{in}^{(k)}\boldsymbol{a}_{in}/\sqrt{\boldsymbol{a}_{in}^{\mathsf{H}}\hat{\boldsymbol{U}}_{in}^{(k)}\boldsymbol{a}_{in}}, \tag{16}$$

$$\boldsymbol{\zeta}_{in} \leftarrow \boldsymbol{u}_{in} - \boldsymbol{w}_{in}, \tag{17}$$

$$\boldsymbol{W}_i \leftarrow \boldsymbol{W}_i + \boldsymbol{e}_n\boldsymbol{\zeta}_{in}^{\mathsf{H}}, \tag{18}$$

$$\boldsymbol{A}_i \leftarrow \left(\boldsymbol{E}_N - \frac{\boldsymbol{a}_{in}\boldsymbol{\zeta}_{in}^{\mathsf{H}}}{1 + \boldsymbol{\zeta}_{in}^{\mathsf{H}}\boldsymbol{a}_{in}}\right)\boldsymbol{A}_i, \tag{19}$$

where $\hat{\boldsymbol{U}}_{in}^{(0)}$ is initialized as $\varepsilon^{-1}\boldsymbol{E}_N$.

III. Proposed method

A. Motivation

The update rule of the conventional OIVA is heuristically derived from that of offline AuxIVA, and the corresponding optimization problem has not been discussed. Although the conventional OIVA has been experimentally shown to perform well, its theoretical background remains ambiguous.

In this paper, to discuss the theoretical background of OIVA, we introduce a framewise probabilistic generative model. In offline IVA, the separated signals are assumed to be independently and identically distributed across time frames; thus, the first and last frames of an observed signal are treated with equal weight. In an online scenario, new frames of an observed signal arrive continuously, and we must consider a very long observed signal. In this case, it is unrealistic to assume that the spatial characteristics are stationary. If we simply apply offline IVA to such a long observed signal, its early frames can affect the separation performance in the current frame. Therefore, it is desirable to treat the recent frames with larger weights than the early frames. To achieve this, we design a probabilistic generative model that changes with each frame.

Then, for maximum likelihood estimation based on the proposed framewise probabilistic generative model, we derive the update rules suitable for online scenarios using IP and ISS, and also discuss the relationship between the proposed and conventional OIVAs.

B. Framewise probabilistic generative model

When the observed signal up to the kth frame is obtained, (3) is represented as

$$-\sum_n \log p(\{y_{ijn}\}_{i,j}^{I,k}) - k\sum_i \log|\det \boldsymbol{W}_i|^2. \tag{20}$$

To weight each frame of the observed signal, we modify the scale matrix of the spherical multivariate Laplace distribution with a frame-dependent weighted identity matrix as

$$p\Big(\{y_{ijn}\}_i^I; \boldsymbol{0}_I, \big(\rho_j^{(k)}\big)^{-2}\boldsymbol{E}_I\Big) \propto \exp\left(-\rho_j^{(k)}\sqrt{\sum_i |y_{ijn}|^2}\right), \tag{21}$$

where $\rho_j^{(k)} \geq 0$ is a weight parameter of the jth frame when the observed signal up to the kth frame is obtained. Here, the autoregressive update rule (13) corresponds to exponential smoothing. Then, we set $\rho_j^{(k)}$ to be

$$\rho_j^{(k)} \propto \beta^{k-j}, \tag{22}$$

where $\beta \in [0, 1)$ denotes the forgetting factor for the proposed model. Additionally, to determine the scale of the weight parameter $\rho_j^{(k)}$, we impose the following constraint:

$$\frac{1}{k}\sum_{j=1}^{k}\rho_j^{(k)} = 1. \tag{23}$$

From (22) and (23), we obtain

$$\rho_j^{(k)} = k\frac{1-\beta}{1-\beta^k}\beta^{k-j}. \tag{24}$$

By substituting (21) into (20), the proposed cost function of IVA at the kth frame $\mathcal{L}_{\mathrm{IVA}}^{(k)}$ is defined as

$$\mathcal{L}_{\mathrm{IVA}}^{(k)} = \sum_{j=1}^{k}\rho_j^{(k)}\sum_n\sqrt{\sum_i|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2} - k\sum_i\log|\det\boldsymbol{W}_i|^2. \tag{25}$$

Note that since the proposed cost function $\mathcal{L}_{\mathrm{IVA}}^{(k)}$ converges to the cost function of offline IVA $\mathcal{L}_{\mathrm{IVA}}$ when $k = J$ and $\beta \to 1$, it can comprehensively represent both offline and online IVAs.

C. Update rules for proposed OIVA

Next, we derive the update rule for the proposed cost function $\mathcal{L}_{\mathrm{IVA}}^{(k)}$. By utilizing the relationship between a concave function and its tangent line, we consider the following auxiliary function of $\mathcal{L}_{\mathrm{IVA}}^{(k)}$:

$$\begin{aligned}\tilde{\mathcal{L}}_{\mathrm{IVA}}^{(k)} =& \sum_{j=1}^{k}\rho_j^{(k)}\sum_n\left(\frac{\sum_i|\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ij}|^2}{2r_{jn}} + \frac{r_{jn}}{2}\right)\\ &- k\sum_i\log|\det\boldsymbol{W}_i|^2\\ =& k\left(\sum_{i,n}\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{\Phi}_{in}^{(k)}\boldsymbol{w}_{in} - \sum_i\log|\det\boldsymbol{W}_i|^2\right) + \text{const.},\end{aligned} \tag{26}$$

where $\boldsymbol{\Phi}_{in}^{(k)}$ is a frame-dependent-weighted covariance matrix of the observed signal and defined as

$$\boldsymbol{\Phi}_{in}^{(k)} = \frac{1}{k}\sum_{j=1}^{k}\rho_j^{(k)}\frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{2r_{jn}} = \frac{1-\beta}{1-\beta^k}\sum_{j=1}^{k}\beta^{k-j}\frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{2r_{jn}}, \tag{27}$$

and $\tilde{\mathcal{L}}_{\mathrm{IVA}}^{(k)} = \mathcal{L}_{\mathrm{IVA}}^{(k)}$ holds if and only if (8) holds. We can minimize $\mathcal{L}_{\mathrm{IVA}}^{(k)}$ by iteratively repeating the following two processes:

- Minimizing $\tilde{\mathcal{L}}_{\mathrm{IVA}}^{(k)}$ with respect to $\boldsymbol{W}_i$
- Updating r_{jn} and $\boldsymbol{\Phi}_{in}^{(k)}$ using (8) and (27), respectively

For minimizing the auxiliary function $\tilde{\mathcal{L}}_{\mathrm{IVA}}^{(k)}$ with respect to the demixing matrix $\boldsymbol{W}_i$, both IP and ISS can be applied. It is guaranteed that both the update rules based on IP and ISS monotonically nonincrease the cost function $\mathcal{L}_{\mathrm{IVA}}^{(k)}$.

For real-time execution, the computational cost at each frame must be reduced. Unfortunately, the auxiliary variable and the frame-dependent-weighted covariance matrix need to be recalculated each time that the demixing matrix is updated, and the computational cost increases as the number of frames k increases. Thus, we first transform (27) as

$$\begin{aligned}\boldsymbol{\Phi}_{in}^{(k)} =& \frac{1-\beta}{1-\beta^k}\frac{\boldsymbol{x}_{ik}\boldsymbol{x}_{ik}^{\mathsf{H}}}{2r_{kn}} \\ &+ \frac{\beta-\beta^k}{1-\beta^k}\left(\frac{1-\beta}{1-\beta^{k-1}}\sum_{j=1}^{k-1}\beta^{(k-1)-j}\frac{\boldsymbol{x}_{ij}\boldsymbol{x}_{ij}^{\mathsf{H}}}{2r_{jn}}\right). \quad (28)\end{aligned}$$

To reduce the computational cost, we fix the auxiliary variables r_{jn} $(j = 1, ..., (k-1))$, resulting in the representation of the second term of the right-hand side of (28) using the estimate of the $(k-1)$th frame-dependent-weighted covariance matrix $\boldsymbol{\Phi}_{in}^{(k-1)}$, $\hat{\boldsymbol{\Phi}}_{in}^{(k-1)}$. As a result, the update of $\hat{\boldsymbol{\Phi}}_{in}^{(k)}$ can be expressed as

$$\hat{\boldsymbol{\Phi}}_{in}^{(k)} = \frac{1-\beta}{1-\beta^k}\frac{\boldsymbol{x}_{ik}\boldsymbol{x}_{ik}^{\mathsf{H}}}{2r_{kn}} + \frac{\beta-\beta^k}{1-\beta^k}\hat{\boldsymbol{\Phi}}_{in}^{(k-1)}, \quad (29)$$

where $r_{kn} = \sqrt{\sum_i |\boldsymbol{w}_{in}^{\mathsf{H}}\boldsymbol{x}_{ik}|^2}$ and $\hat{\boldsymbol{\Phi}}_{in}^{(1)}$ is set to $\boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^{\mathsf{H}}/2r_{1n}+\varepsilon\boldsymbol{E}_N$ for stability. Finally, the update rule of the proposed OIVA using IP and the Sherman–Morrison formula is expressed as

$$\boldsymbol{\nu}_{in} \leftarrow \hat{\boldsymbol{\Psi}}_{in}^{(k-1)}\boldsymbol{x}_{ik}, \quad (30)$$

$$\hat{\boldsymbol{\Psi}}_{in}^{(k)} \leftarrow \frac{1-\beta^k}{\beta-\beta^k}\left(\hat{\boldsymbol{\Psi}}_{in}^{(k-1)} - \frac{\boldsymbol{\nu}_{in}\boldsymbol{\nu}_{in}^{\mathsf{H}}}{\boldsymbol{x}_{ik}^{\mathsf{H}}\boldsymbol{\nu}_{in} + \frac{2r_{kn}(\beta-\beta^k)}{(1-\beta)}}\right), \quad (31)$$

$$\boldsymbol{\xi}_{in} \leftarrow \hat{\boldsymbol{\Psi}}_{in}^{(k)}\boldsymbol{a}_{in}/\sqrt{\boldsymbol{a}_{in}^{\mathsf{H}}\hat{\boldsymbol{\Psi}}_{in}^{(k)}\boldsymbol{a}_{in}}, \quad (32)$$

$$\boldsymbol{\varphi}_{in} \leftarrow \boldsymbol{\xi}_{in} - \boldsymbol{w}_{in}, \quad (33)$$

$$\boldsymbol{W}_i \leftarrow \boldsymbol{W}_i + \boldsymbol{e}_n\boldsymbol{\varphi}_{in}^{\mathsf{H}}, \quad (34)$$

$$\boldsymbol{A}_i \leftarrow \left(\boldsymbol{E}_N - \frac{\boldsymbol{a}_{in}\boldsymbol{\varphi}_{in}^{\mathsf{H}}}{1+\boldsymbol{\varphi}_{in}^{\mathsf{H}}\boldsymbol{a}_{in}}\right)\boldsymbol{A}_i, \quad (35)$$

where $\hat{\boldsymbol{\Psi}}_{in}^{(1)}$ is set to $\varepsilon^{-1}\boldsymbol{E}_N - \boldsymbol{x}_{i1}\boldsymbol{x}_{i1}^{\mathsf{H}}/(2r_{1n}\varepsilon^2+\varepsilon\boldsymbol{x}_{i1}^{\mathsf{H}}\boldsymbol{x}_{i1})$ for stability. On the other hand, the update rule of the proposed OIVA using ISS is expressed as

$$q_{inn'} \leftarrow \begin{cases} \dfrac{\boldsymbol{w}_{in'}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{in'}^{(k)}\boldsymbol{w}_{in}}{\boldsymbol{w}_{in}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{in'}^{(k)}\boldsymbol{w}_{in}}, & (\text{if } n' \neq n) \\ 1 - (\boldsymbol{w}_{in}^{\mathsf{H}}\hat{\boldsymbol{\Phi}}_{in}^{(k)}\boldsymbol{w}_{in})^{-\frac{1}{2}}, & (\text{if } n' = n) \end{cases} \quad (36)$$

$$\boldsymbol{W}_i \leftarrow \boldsymbol{W}_i - (q_{in1}, ..., q_{inN})^{\mathsf{T}}\boldsymbol{w}_{in}^{\mathsf{H}}. \quad (37)$$

D. Relationship between proposed and conventional OIVAs

When $\alpha = \beta$ (< 1) and $k \to \infty$, it can be seen that the update rule of the frame-dependent-weighted covariance matrix of the observed signal (29) converges to the update rule of the weighted covariance matrix in the conventional OIVAs (7). Therefore, the conventional OIVAs can be interpreted as the limiting case of the proposed OIVAs where an infinitely long time has passed. On the other hand, when k is small, the proposed OIVAs assign less weight to the previous estimate of the weighted covariance matrix of the observed signal than the conventional OIVAs. In such situations, since it is difficult to obtain this weighted covariance matrix accurately, the proposed OIVAs are expected to quickly capture the spatial characteristics by assigning a large weight to the latest observed signal.

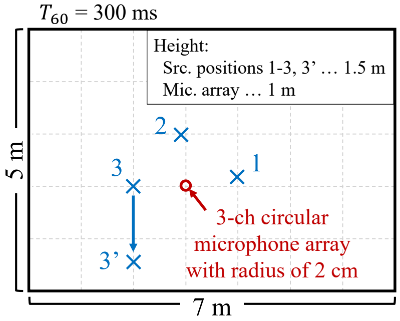
T_{60} = 300 ms
Height:
Src. positions 1-3, 3' … 1.5 m
Mic. array … 1 m
5 m
7 m
1
2
3
3'
3-ch circular microphone array with radius of 2 cm

Fig. 1. Room layout for simulating impulse response. In experiment (i), all sources are stationary at positions 1–3. In experiment (ii), two sources at positions 1 and 2 are stationary and one source at position 3 moves to position 3' between 10 and 20 s after start.

IV. Experiments

A. Experimental conditions

We conducted two simulation experiments: (i) a case where all sources are stationary and (ii) a case where one of the sources moves. We used speech signals from the JNAS dataset [10] and chose 10 sets of three speakers. For each speaker, we created a dry source by concatenating speech signals with a total length of 60 s. All sources were convolved with the room impulse response generated using the image source method implemented in Pyroomacoustics [11] and then mixed so that each convolved signal had equal power. Fig. 1 shows the room layout for simulation. The reverberation time T_{60} was set to 300 ms. In experiment (i), all sources remained stationary at positions 1–3 shown in Fig. 1. In experiment (ii), two sources were stationary at positions 1 and 2, and one source moved at a constant speed from position 3 to position 3' between 10 and 20 s after the start. The sampling rate was 16 kHz.

We compared four methods: *Conv-IP* and *Conv-ISS* are the conventional OIVAs using IP and ISS, respectively; *Prop-IP* and *Prop-ISS* are the proposed OIVAs using IP and ISS, respectively. For both experiments, we set the forgetting parameters α and β to 0.99, the number of iterations per frame to 2, and a small constant for stability ε to 10^{-3}. STFT was performed using a 64-ms-long Hamming window with a shift length of 32 ms. The implementation was carried out using Python on a PC equipped with Intel Core i9-13900KF CPU and 128 GB of RAM. Only the CPU was used for computations.

To evaluate the real-time source separation performance, we first calculated the segmentwise source-to-distortion ratio

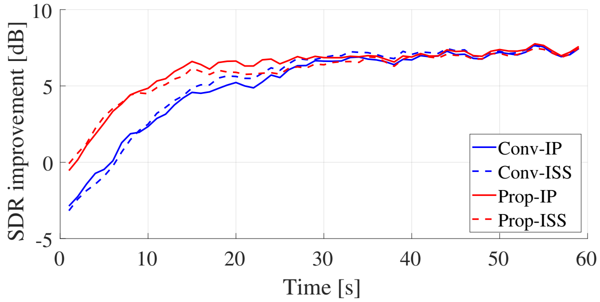

Fig. 2. Average SDR improvements for each method in experiment (i).

TABLE I
AVERAGE AND MAXIMUM PROCESSING TIMES PER FRAME FOR EACH METHOD IN EXPERIMENT (I).

Method	Conv-IP	Conv-ISS	Prop-IP	Prop-ISS
Ave. [ms]	1.86	2.00	1.86	2.01
Max. [ms]	2.29	4.11	3.71	3.99

(SDR) [12] improvement for each source using the signals between $(l-1)$ s and $(l+1)$ s $(l=1,\ldots,59)$. In experiment (i), we used the average SDR improvement across all 30 sources. In experiment (ii), we used the average SDR improvement across 20 stationary sources and that across 10 moving sources. Additionally, we calculated the average and maximum processing times per frame for each method in experiment (i).

B. Online source separation performance

Fig. 2 shows the average SDR improvements for each method in experiment (i). We confirmed that the separation performance of the proposed methods converges faster than that of the conventional methods. Furthermore, after sufficient time had passed, the separation performance of all the methods became almost the same. Table 1 shows the average and maximum processing times per frame in experiment (i). Since the maximum processing time of all the methods was significantly lower than the frame length (32 ms), it was confirmed that all the methods can operate in real time. Additionally, when using the same demixing matrix estimation method, there was almost no difference in the average processing time between the conventional and proposed methods. These results indicate that the proposed methods achieve faster convergence than the conventional methods without any performance degradation.

Fig. 3 shows the average SDR improvement of stationary sources and that of moving sources for each method in experiment (ii). As shown in this result, we confirmed that both the conventional and proposed methods perform well even in situations where a moving source exists.

V. Conclusion

In this paper, we proposed an optimization problem by introducing a framewise probabilistic generative model for OIVA and derived the update rule. The theoretical background of the conventional OIVA is also derived as the limiting case of the proposed OIVA. Through experiments, we confirmed that the separation performance of the proposed methods converges faster than that of the conventional methods and that the proposed methods perform well even in situations where a moving source exists.

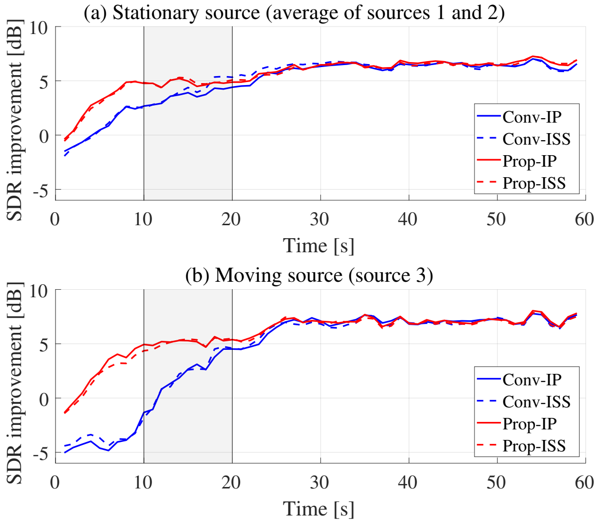

Fig. 3. Average SDR improvements of stationary sources (top panel) and moving sources (bottom panel) for each method in experiment (ii). Gray shaded area represents time interval in which source at position 3 moved to position 3'.

Acknowledgment

The authors would like to thank Assistant Professor Taishi Nakashima and Professor Nubutaka Ono (Tokyo Metropolitan University) for providing the experimental codes of conventional online IVAs.

References

[1] H. Sawada et al., "A review of blind source separation methods: Two converging routes to ILRMA originating from ICA and NMF," *APSIPA TSIP*, vol. 8, no. e12, pp. 1–14, 2019.

[2] T. Kim et al., "Blind source separation exploiting higher-order frequency dependencies," *IEEE TASLP*, vol. 15, no. 1, pp. 70–79, 2007.

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

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

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

[6] T. Taniguchi et al., "An auxiliary-function approach to online independent vector analysis for real-time blind source separation," in *Proc. HSCMA*, 2014, pp. 107–111.

[7] T. Nakashima and N. Ono, "Inverse-free online independent vector analysis with flexible iterative source steering," in *Proc. APSIPA ASC*, 2022, pp. 750–754.

[8] R. Scheibler and N. Ono, "Fast and stable blind source separation with rank-1 updates," in *Proc. ICASSP*, 2020, pp. 236–240.

[9] N. Murata et al., "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 2001, pp. 1–24, 2001.

[10] K. Itou et al., "JNAS: Japanese speech corpus for large vocabulary continuous speech recognition research," *JASJ*, vol. 20, no. 3, pp. 199–206, 1999.

[11] R. Scheibler et al., "Pyroomacoustics: A python package for audio room simulation and array processing algorithms," in *Proc. ICASSP*, 2018, pp. 351–355.

[12] E. Vincent et al., "Performance measurement in blind audio source separation," *IEEE TASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.