Prior-Guided Source Separation with Direct Update of Back-projected Demixing Vectors

Kukuru Koiso, Taishi Nakashima, Nobutaka Ono*
Tokyo Metrepolitan University, Tokyo, Japan
E-mail: koiso-kukuru@ed.tmu.ac.jp, taishi@tmu.ac.jp, onono@tmu.ac.jp

***Abstract*—In this study, we propose a blind source separation (BSS) method that incorporates prior knowledge of demixing vectors by directly optimizing their back-projected versions. In conventional BSS frameworks, the demixing matrix is first optimized under scale ambiguity and then rescaled through back-projection, which yields the actual demixing matrix used for separation. While some existing approaches have utilized prior information as regularization in the optimization of demixing vectors, a mismatch in scale between the optimized vectors and the prior vectors, which are often defined in the back-projected domain, makes it difficult to control the strength of the prior influence, potentially degrading performance. To address this issue, we build upon our recent formulation that reparameterizes the optimization to directly update the back-projected demixing matrix. By incorporating the prior information in this domain, our method allows for more consistent and effective use of prior knowledge. Simulated experiments demonstrate that the proposed method achieves superior separation performance compared to conventional approaches.**

I. Introduction

Source separation is a technique of estimating the original source signals from observed signals that are mixtures of multiple sources. Among various source separation techniques, Blind Source Separation (BSS) is a method that separates individual source signals using only the observed mixtures, without any prior information such as the positions of the sources or microphones. A commonly used BSS method is Independent Vector Analysis (IVA) [1], [2], which assumes the co-occurrence of frequency components belonging to the same source. An extension of IVA, known as Auxiliary-function-based Independent Vector Analysis (AuxIVA) [3], has also been proposed, which incorporates a majorization-minimization algorithm [4]. In addition, recent studies have incorporated prior knowledge into BSS frameworks to improve separation performance [5].

One known issue in IVA-based BSS methods is the ambiguity in the scale and phase of the separated signals. To resolve this ambiguity, a post-processing called projection back [6] is commonly applied, yielding a rescaled demixing matrix which we call a "back-projected" demixing matrix. However, since the demixing matrix actually used for separation is obtained after projection back, conventional approaches have introduced constraints during the optimization of the pre-projection demixing matrix. In contrast, our recent study [7] re-defined the independent variables using the conditions satisfied by the back-projected demixing matrix, which enables direct optimization of the demixing matrix before projection back. In this study, we propose direct incorporating prior information into the optimization of the back-projected demixing matrix. To evaluate its effectiveness, simulation experiments were conducted to compare its separation performance with that of conventional methods.

II. Problem Formulation

We assume that the number of sources and microphones is equal to K. In the Short-Time Fourier Transform (STFT) domain, let k, f, and t denote the indices of the source, frequency bin, and time frame, respectively. The source signals are represented as $\boldsymbol{s}_{ft} = \begin{pmatrix} s_{1ft} & \dots & s_{Kft} \end{pmatrix}^\top \in \mathbb{C}^K$ and the observed signals $\boldsymbol{x}_{ft} \in \mathbb{C}^K$ are assumed to follow the mixing model:

$$\boldsymbol{x}_{ft} = \sum_{k=1}^{K} \boldsymbol{a}_{kf} s_{kft} := \boldsymbol{A}_f \boldsymbol{s}_{ft}, \tag{1}$$

where, the indices k, f, and t range over $(1, \dots, K)$, $(1, \dots, F)$, and $(1, \dots, T)$, respectively. The vector $\boldsymbol{a}_{kf} := \begin{pmatrix} a_{1kf} & \dots & a_{Kkf} \end{pmatrix}^\top \in \mathbb{C}^K$ is the steering vector from source s_{kft} to each microphone. The matrix $\boldsymbol{A}_f := \begin{pmatrix} \boldsymbol{a}_{1f} & \dots & \boldsymbol{a}_{Kf} \end{pmatrix} \in \mathbb{C}^{K \times K}$ represents the mixing matrix.

In BSS with equal numbers of sources and microphones, the estimated source signals are obtained by estimating the inverse of the mixing model:

$$\boldsymbol{y}_{ft} = \boldsymbol{W}_f \boldsymbol{x}_{ft}, \tag{2}$$

where $\boldsymbol{y}_{ft}$ denotes the separated signals. The matrix $\boldsymbol{W}_f := \begin{pmatrix} \boldsymbol{w}_{1f} & \dots & \boldsymbol{w}_{Kf} \end{pmatrix}^H \in \mathbb{C}^{K \times K}$ is called the demixing matrix, and $\boldsymbol{w}_{kf} \in \mathbb{C}^K$ is called the demixing vector (row vector of $\boldsymbol{W}_f$). Ideally, $\boldsymbol{W}_f$ equals $\boldsymbol{A}_f^{-1}$.

In this study, we consider the problem of source separation assuming that an approximate back-projected demixing matrix is available. Such prior information can be obtained, for example, from DOA estimation, pre-trained dictionaries tailored to specific recording environments, or past estimates in online processing. While these estimates are typically noisy or inaccurate, they can still serve as useful priors to guide the separation process. Building on this assumption, we formulate the optimization problem of the demixing matrix by incorporating constraint terms based on prior information about the back-projected matrix.

979-8-3315-7206-8/25/$31.00 ©2025 IEEE

III. Conventional Methods

A. Demixing Matrix Optimization

In this study, we consider BSS with a time-varying Gaussian source model. Each source s_{kft} is assumed to independently follow a complex Gaussian distribution with zero mean and a variance r_{kt} that is common across frequency bins f for each time frame t. Under these assumptions, the cost function of BSS is given by:

$$J(\mathcal{W}, R) = \frac{1}{T}\sum_{f=1}^{F}\sum_{t=1}^{T}\left(\sum_{k=1}^{K}\left(\frac{|\boldsymbol{w}_{kf}^H\boldsymbol{x}_{ft}|^2}{r_{kt}} + \log r_{kt}\right) - 2\log|\det \boldsymbol{W}_f|\right), \quad (3)$$

where $\boldsymbol{R}$ denotes the set of r_{kt} for all k and t, and $\mathcal{W}$ denotes the set of $\boldsymbol{W}_f$ for all f. The goal is to estimate the parameters $\mathcal{W}$ and $\boldsymbol{R}$ by minimizing (3) [8].

B. BSS with Spatial Regularization

As an approach to incorporating prior information into BSS, Mitsui et al. [5] proposed a method that introduces a spatial regularization term into the framework of Independent Low-Rank Matrix Analysis (ILRMA) [5]. In their method, prior demixing vectors are used to guide the estimated demixing vectors toward the desired spatial characteristics, thereby improving the separation performance. This concept is not limited to ILRMA and can also be applied to other BSS frameworks such as IVA.

In this work, we adapt this idea to AuxIVA by introducing a penalty term into its cost function. This penalty encourages each demixing vector to be close to the corresponding prior vector$\hat{\boldsymbol{w}}_{kf}$. The formulation is expressed as follows:

$$J_{\text{PG}} = J_{\text{AuxIVA}} + \sum_{f=1}^{F}\sum_{k=1}^{K}\lambda_k\|\boldsymbol{w}_{kf} - \hat{\boldsymbol{w}}_{kf}\|^2. \quad (4)$$

where J_{AuxIVA} denote the original cost function of AuxIVA [3], and λ_k is the weight for the regularization term. We refer to this formulation as Prior-Guided AuxIVA (PG-AuxIVA) in the rest of this paper. Although such use of prior demixing vectors has been referred to as "spatial regularization" in previous studies, we note that the term "regularization" typically implies indirect constraints, such as sparsity or smoothness. In contrast, the approach in both the previous work and ours directly guides the solution toward a given approximate one. Therefore, we refer to it as "prior-guided" in this paper.

C. Projection Back

This paper proposes an extension of our previous work [7], which directly updates the back-projected demixing matrix. We henceforth refer to this method as Back-Projected AuxIVA (BP-AuxIVA) throughout this paper. First, we describe the previous method in detail. In general, the separated signals in BSS suffer from scale ambiguity. Projection back is a post-processing step to resolve this scale ambiguity after the demixing matrix is estimated [6]. Specifically, let ℓ be the index of the reference microphone, then the back-projected demixing matrix is defined as:

$$\tilde{\boldsymbol{W}}_f = \text{diag}(a_{\ell 1f}, a_{\ell 2f}, \cdots, a_{\ell Kf})\boldsymbol{W}_f, \quad (5)$$

$$\boldsymbol{y}_{ft} = \tilde{\boldsymbol{W}}_f\boldsymbol{x}_{ft}, \quad (6)$$

where $\text{diag}(\cdot)$ denotes a diagonal matrix whose diagonal elements are the listed components. Here, the coefficients $a_{\ell 1f}, \ldots, a_{\ell Kf}$ are elements from the ℓ-th row of the inverse matrix $\boldsymbol{W}_f^{-1}$. This allows the separated signals to have the same scale as the observation at the reference microphone. In what follows, we use a tilde to denote the demixing matrix after projection back. The original demixing matrix before projection back can be written as:

$$\boldsymbol{C}_f := \text{diag}\left(\frac{1}{a_{\ell 1f}}, \frac{1}{a_{\ell 2f}}, \ldots, \frac{1}{a_{\ell Kf}}\right), \quad (7)$$

and

$$\boldsymbol{W}_f = \boldsymbol{C}_f\tilde{\boldsymbol{W}}_f. \quad (8)$$

Moreover, the back-projected demixing matrix satisfies the constraint:

$$\sum_{k=1}^{K}\tilde{\boldsymbol{w}}_k = \boldsymbol{e}_\ell, \quad (9)$$

as shown in [7]. Without loss of generality, we assume $\ell = 1$ in the following discussions.

D. Direct Optimization of the Back-Projected Demixing Matrix

BP-AuxIVA directly optimizes the back-projected demixing matrix. In this approach, $\tilde{\boldsymbol{w}}_{1f}, \cdots, \tilde{\boldsymbol{w}}_{(K-1)f}$ and $\boldsymbol{C}_f = \text{diag}(c_{1f}, \cdots, c_{Kf})$ are treated as independent variables. To ensure that $\tilde{\boldsymbol{W}}_f = (\tilde{\boldsymbol{w}}_{1f}, \tilde{\boldsymbol{w}}_{2f}, \ldots, \tilde{\boldsymbol{w}}_{Kf})^H$ always satisfies the projection-back constraint, the last row vector $\tilde{\boldsymbol{w}}_{Kf}$ is defined as a dependent variable:

$$\tilde{\boldsymbol{w}}_{Kf} = \boldsymbol{e}_1 - \sum_{k=1}^{K-1}\tilde{\boldsymbol{w}}_{kf}. \quad (10)$$

Since $\tilde{\boldsymbol{w}}_{Kf}$ is a dependent variable as defined in (10), it also implicitly includes the variables $\tilde{\boldsymbol{w}}_{kf}$ for $k = 1, \ldots, K-1$ in the Kth row of $\tilde{\boldsymbol{W}}_f$. To organize this dependency, we define the following variables:

$$\tilde{\boldsymbol{W}}_f = \boldsymbol{G}\tilde{\boldsymbol{W}}_f^\dagger.$$

Where, $\tilde{\boldsymbol{W}}_f^\dagger$ is a matrix composed of $\tilde{\boldsymbol{w}}_{1f}, \cdots, \tilde{\boldsymbol{w}}_{(K-1)f}$ and $\boldsymbol{e}_1$. $\boldsymbol{G}$ is a transformation matrix.

In this case, since $\det\boldsymbol{G} = 1$, it follows that $\det\tilde{\boldsymbol{W}}_f = \det\tilde{\boldsymbol{W}}_f^\dagger$. Assuming $\boldsymbol{R}$ has already been optimized, we focus on a specific $\tilde{\boldsymbol{w}}_{kf}$ ($k = 1, 2, \ldots, K-1$), treating the others as fixed. We define:

$$\boldsymbol{b}_{kf} = |c_{Kf}|^2 \boldsymbol{V}_{Kf} \left(\boldsymbol{e}_1 - \sum_{l \neq k, K} \tilde{\boldsymbol{w}}_{lf} \right), \tag{11}$$

$$\boldsymbol{D}_{kf} = |c_{kf}|^2 \boldsymbol{V}_{kf} + |c_{Kf}|^2 \boldsymbol{V}_{Kf}, \tag{12}$$

and ignoring constant terms irrelevant to optimization, the cost function is simplified to:

$$\begin{aligned} J(\tilde{\mathcal{W}}, \mathcal{C}) = &\sum_{f=1}^{F} \sum_{k=1}^{K-1} \left(\tilde{\boldsymbol{w}}_{kf}^H \boldsymbol{D}_{kf} \tilde{\boldsymbol{w}}_{kf} - \boldsymbol{b}_{kf}^H \tilde{\boldsymbol{w}}_{kf} - \tilde{\boldsymbol{w}}_{kf}^H \boldsymbol{b}_{kf} \right) \\ &- \sum_{f=1}^{F} 2 \log |\det \boldsymbol{C}_f \tilde{\boldsymbol{W}}_f^{\dagger}|. \end{aligned} \tag{13}$$

This objective function is the same form as the one appeared in [5]. Therefore, the update rules are given by:

$$\boldsymbol{u}_{kf} = \boldsymbol{D}_{kf}^{-1} (\tilde{\boldsymbol{W}}_f^{\dagger})^{-1} \boldsymbol{e}_k, \tag{14}$$

$$\boldsymbol{q}_{kf} = \boldsymbol{D}_{kf}^{-1} \boldsymbol{b}_{kf}, \tag{15}$$

$$g_{kf} = \boldsymbol{u}_{kf}^H \boldsymbol{D}_{kf} \boldsymbol{u}_{kf}, \tag{16}$$

$$h_{kf} = \boldsymbol{u}_{kf}^H \boldsymbol{D}_{kf} \boldsymbol{q}_{kf}, \tag{17}$$

$$\tilde{\boldsymbol{w}}_{kf} \leftarrow \begin{cases} \frac{\boldsymbol{u}_{kf}}{\sqrt{g_{kf}}} + \boldsymbol{q}_{kf} & (h_{kf} = 0), \\ \frac{h_{kf}}{2 g_{kf}} \left(-1 + \sqrt{1 + \frac{4 g_{kf}}{|h_{kf}|^2}} \right) \boldsymbol{u}_{kf} + \boldsymbol{q}_{kf} & (\text{otherwise}). \end{cases} \tag{18}$$

On the other hand, the optimization of c_{kf} is analogous to the normalization step in IP [3], and is performed as follows:

$$c_{kf} \leftarrow \frac{1}{\sqrt{\boldsymbol{w}_{kf}^H \boldsymbol{V}_{kf} \boldsymbol{w}_{kf}}}. \tag{19}$$

IV. Proposed Method

A. Incorporating Prior Information using Demixing Vectors

In the following, we incorporate the prior information term into (13) using a known back-projected demixing vector $\boldsymbol{z}_{kf}$. Let λ_k be the weighting coefficient that controls the influence of the prior knowledge for each demixing vector, and $\tilde{\boldsymbol{w}}_{kf}$ be the demixing vector to be optimized.This prior term encourages each demixing vector $\tilde{w}_{kf}$ to be close to the corresponding known vector $\boldsymbol{z}_{kf}$, reflecting prior knowledge such as estimated steering vectors or ideal demixing vectors. The incorporated prior information term $P(\tilde{\mathcal{W}})$ is defined as:

$$P(\tilde{\mathcal{W}}) = \sum_{f=1}^{F} \sum_{k=1}^{K} \lambda_k \|\tilde{\boldsymbol{w}}_{kf} - \boldsymbol{z}_{kf}\|^2. \tag{20}$$

In BP-AuxIVA, $\tilde{\boldsymbol{w}}_{Kf}$ is treated not as an independent variable but as a dependent one. Thus, by substituting $\tilde{\boldsymbol{w}}_{Kf} = \boldsymbol{e}_1 - \sum_{k=1}^{K-1} \tilde{\boldsymbol{w}}_{kf}$ into (20), the expression becomes:

$$\begin{aligned} P(\tilde{\mathcal{W}}) = &\sum_{f=1}^{F} \left(\sum_{k=1}^{K-1} \lambda_k \|\tilde{\boldsymbol{w}}_{kf} - \boldsymbol{z}_{kf}\|^2 \right. \\ &\left. + \lambda_K \left\| \left(\boldsymbol{e}_1 - \sum_{k=1}^{K-1} \tilde{\boldsymbol{w}}_{kf} \right) - \boldsymbol{z}_{Kf} \right\|^2 \right). \end{aligned} \tag{21}$$

In BSS algorithms, $\tilde{\boldsymbol{w}}_{kf}$ is typically updated sequentially. Hence, for $m \neq k$ (where $m = 1, \ldots, K-1$), the other vectors are treated as constants, and we rewrite:

$$\sum_{k=1}^{K-1} \tilde{\boldsymbol{w}}_{kf} = \tilde{\boldsymbol{w}}_{kf} + \sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf}. \tag{22}$$

By excluding the terms unrelated to the update of $\tilde{\boldsymbol{w}}_{kf}$, we get:

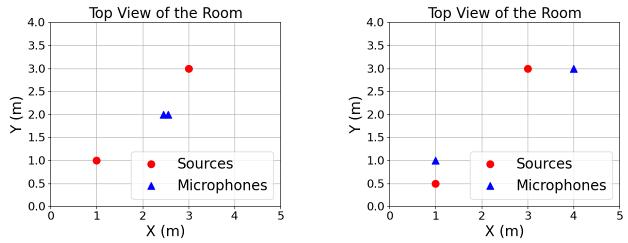
$$\begin{aligned} P(\tilde{\mathcal{W}}) = &\sum_{f=1}^{F} \sum_{k=1}^{K-1} \Big(\lambda_k \big(\tilde{\boldsymbol{w}}_{kf}^H \tilde{\boldsymbol{w}}_{kf} - \boldsymbol{z}_{kf}^H \tilde{\boldsymbol{w}}_{kf} - \tilde{\boldsymbol{w}}_{kf}^H \boldsymbol{z}_{kf} \big) \\ &+ \lambda_K \big(- \boldsymbol{e}_1^H \tilde{\boldsymbol{w}}_{kf} - \tilde{\boldsymbol{w}}_{kf}^H \boldsymbol{e}_1 \\ &+ \tilde{\boldsymbol{w}}_{kf}^H (\sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf}) + (\sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf})^H \tilde{\boldsymbol{w}}_{kf} \\ &+ \tilde{\boldsymbol{w}}_{kf}^H \tilde{\boldsymbol{w}}_{kf} + \boldsymbol{z}_{Kf}^H \tilde{\boldsymbol{w}}_{kf} + \tilde{\boldsymbol{w}}_{kf}^H \boldsymbol{z}_{Kf} \big) \Big). \end{aligned} \tag{23}$$

Grouping the terms by $\tilde{\boldsymbol{w}}_{kf}^H \tilde{\boldsymbol{w}}_{kf}$, $\tilde{\boldsymbol{w}}_{kf}$, and $\tilde{\boldsymbol{w}}_{kf}^H$, we obtain:

$$\begin{aligned} P(\tilde{\mathcal{W}}) = &\sum_{f=1}^{F} \sum_{k=1}^{K-1} \Big(\tilde{\boldsymbol{w}}_{kf}^H (\lambda_k + \lambda_K) \boldsymbol{I}_{K,K} \tilde{\boldsymbol{w}}_{kf} \\ &- \big(\lambda_k \boldsymbol{z}_{kf}^H + \lambda_K (\boldsymbol{e}_1^H - (\sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf})^H - \boldsymbol{z}_{Kf}^H) \big) \tilde{\boldsymbol{w}}_{kf} \\ &- \tilde{\boldsymbol{w}}_{kf}^H \big(\lambda_k \boldsymbol{z}_{kf} + \lambda_K (\boldsymbol{e}_1 - (\sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf}) - \boldsymbol{z}_{Kf}) \big) \Big), \end{aligned} \tag{24}$$

where, $\boldsymbol{I}_{K,K}$ is the identity matrix of size K. Substituting Eq. (24) into Eq. (13), we obtain:

$$\begin{aligned} &J(\tilde{\mathcal{W}}, \mathcal{C}) \\ &= \sum_{f=1}^{F} \sum_{k=1}^{K-1} \Bigg(\tilde{\boldsymbol{w}}_{kf}^H \big(\boldsymbol{D}_{kf} + (\lambda_k + \lambda_K) \boldsymbol{I}_{K,K} \big) \tilde{\boldsymbol{w}}_{kf} \\ &- \left(\lambda_k \boldsymbol{z}_{kf} + \lambda_K \Big(\boldsymbol{e}_1 - \sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf} - \boldsymbol{z}_{Kf} \Big) \right)^H \tilde{\boldsymbol{w}}_{kf} \\ &- \tilde{\boldsymbol{w}}_{kf}^H \Big(\lambda_k \boldsymbol{z}_{kf} + \lambda_K (\boldsymbol{e}_1 - \sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf} - \boldsymbol{z}_{Kf}) \Big) \Bigg) \\ &- \sum_{f=1}^{F} 2 \log \left| \det \boldsymbol{C}_f \tilde{\boldsymbol{W}}_f^{\dagger} \right|. \end{aligned} \tag{25}$$



(a) Centralized configuration (b) Distributed configuration

Fig. 1: Source and microphone configurations

Thus, by redefining the coefficients as:

$$\widehat{\boldsymbol{D}}_{kf} = \boldsymbol{D}_{kf} + (\lambda_k + \lambda_K)\boldsymbol{I}_{K,K}, \tag{26}$$

$$\widehat{\boldsymbol{b}}_{kf} = \boldsymbol{b}_{kf} + \lambda_k \boldsymbol{z}_{kf} + \lambda_K \left(\boldsymbol{e}_1 - \sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf} - \boldsymbol{z}_{Kf} \right), \tag{27}$$

the cost function takes the following form:

$$J(\tilde{\mathcal{W}}, \mathcal{C}) = \sum_{f=1}^{F} \sum_{k=1}^{K-1} (\tilde{\boldsymbol{w}}_{kf}^H \widehat{\boldsymbol{D}}_{kf} \tilde{\boldsymbol{w}}_{kf} - \widehat{\boldsymbol{b}}_{kf}^H \tilde{\boldsymbol{w}}_{kf} - \tilde{\boldsymbol{w}}_{kf}^H \widehat{\boldsymbol{b}}_{kf}) - \sum_{f=1}^{F} 2 \log |\det \boldsymbol{C}_f \tilde{\boldsymbol{W}}_f^{\dagger}|, \tag{28}$$

which retains the same form as Eq. (13). The update of each variable can be performed in the same manner as Eqs. (14)–(19). The algorithm is shown in Algorithm 1.

Algorithm 1 Prior-Guided Back-Projected AuxIVA

for $i = 1, \ldots, N_\mathrm{i}$ **do**
  for $k = 1, \ldots, K-1$ **do**
    $r_{kt} \leftarrow \sum_{f=1}^{F} |\boldsymbol{w}_{kf}^H \boldsymbol{x}_{ft}|^2$
    $r_{Kt} \leftarrow \sum_{f=1}^{F} |\boldsymbol{w}_{Kf}^H \boldsymbol{x}_{ft}|^2$
    for $f = 1, \ldots, F$ **do**
      $\boldsymbol{V}_{kf} \leftarrow \frac{1}{T} \sum_{t=1}^{T} \frac{\boldsymbol{x}_{ft} \boldsymbol{x}_{ft}^H}{r_{kt}}$
      $\boldsymbol{V}_{Kf} \leftarrow \frac{1}{T} \sum_{t=1}^{T} \frac{\boldsymbol{x}_{ft} \boldsymbol{x}_{ft}^H}{r_{Kt}}$
      $\boldsymbol{D}_{kf} \leftarrow |c_{kf}|^2 \boldsymbol{V}_{kf} + |c_{Kf}|^2 \boldsymbol{V}_{Kf}$
      $\widehat{\boldsymbol{D}}_{kf} \leftarrow \boldsymbol{D}_{kf} + (\lambda_k + \lambda_K)\boldsymbol{I}_{K,K}$
      $\boldsymbol{b}_{kf} \leftarrow |c_{Kf}|^2 \boldsymbol{V}_{Kf} \left(\boldsymbol{e}_1 - \sum_{\ell \neq k, K} \tilde{\boldsymbol{w}}_\ell \right)$
      $\widehat{\boldsymbol{b}}_{kf} \leftarrow \boldsymbol{b}_{kf} + \lambda_k \boldsymbol{z}_{kf} + \lambda_K (\boldsymbol{e}_1 - \sum_{m \neq k} \tilde{\boldsymbol{w}}_{mf} - \boldsymbol{z}_{Kf})$
      $\tilde{\boldsymbol{W}}_f^{\dagger} = \begin{pmatrix} \tilde{\boldsymbol{w}}_{1f} & \ldots & \tilde{\boldsymbol{w}}_{(K-1)f} & \boldsymbol{e}_1 \end{pmatrix}^H$
      $\widehat{\boldsymbol{u}}_{kf} \leftarrow \widehat{\boldsymbol{D}}_{kf}^{-1} (\tilde{\boldsymbol{W}}_f^{\dagger})^{-1} \boldsymbol{e}_k$
      $\widehat{\boldsymbol{q}}_{kf} \leftarrow \widehat{\boldsymbol{D}}_{kf}^{-1} \widehat{\boldsymbol{b}}_{kf}$
      $\widehat{g}_{kf} \leftarrow \widehat{\boldsymbol{u}}_{kf}^H \widehat{\boldsymbol{D}}_{kf} \widehat{\boldsymbol{u}}_{kf}$
      $\widehat{h}_{kf} \leftarrow \widehat{\boldsymbol{u}}_{kf}^H \widehat{\boldsymbol{D}}_{kf} \widehat{\boldsymbol{q}}_{kf}$
      $\tilde{\boldsymbol{w}}_{kf} \leftarrow \frac{2 \widehat{\boldsymbol{u}}_{kf}}{\sqrt{|\widehat{h}_{kf}|^2 + 4 \widehat{g}_{kf}} + |\widehat{h}_{kf}|} + \widehat{\boldsymbol{q}}_{kf}$
      $\tilde{\boldsymbol{w}}_{Kf} \leftarrow \boldsymbol{e}_1 - \sum_{\ell \neq K} \tilde{\boldsymbol{w}}_\ell$
      $c_{kf} \leftarrow \frac{1}{\sqrt{\tilde{\boldsymbol{w}}_{kf}^H \boldsymbol{V}_{kf} \tilde{\boldsymbol{w}}_{kf}}}$
      $c_{Kf} \leftarrow \frac{1}{\sqrt{\tilde{\boldsymbol{w}}_{Kf}^H \boldsymbol{V}_{Kf} \tilde{\boldsymbol{w}}_{Kf}}}$
    end for
  end for
end for

V. Experiments

A. Setup

We evaluate the efficacy of the proposed method through simulated experiments. The sampling rate was set to 16 kHz. The Short-Time Fourier Transform (STFT) used a window length of 8192 samples with a frame shift of 4096 samples and employed a Hann window. As source signals, we used the Japanese Versatile Speech (JVS) corpus [9]. The number of sources and microphones was set to two. For each of the two array configurations, 100 pairs of 20-second audio clips were generated. The reverberation time was set to 400 ms. The spatial arrangements of the sources and microphones are shown in Fig. 1.

The number of iterations for updating the demixing matrix N_i was fixed at 50 for all methods. As prior information, we used back-projected demixing vectors derived from an ideal demixing matrix. Specifically, for each k-th source, we computed the dominant eigenvector of the spatial covariance matrix constructed from the observed signal $\boldsymbol{d}_{kft} := \boldsymbol{a}_{kf} s_{kft}$, which contains only the k-th source signal s_{kft}. This eigenvector was then used as the steering vector $\boldsymbol{a}_{kf}$:

$$\left(\frac{1}{T} \sum_{t=1}^{T} \boldsymbol{d}_{kft} \boldsymbol{d}_{kft}^H \right) \boldsymbol{v} = \lambda \boldsymbol{v}, \tag{29}$$

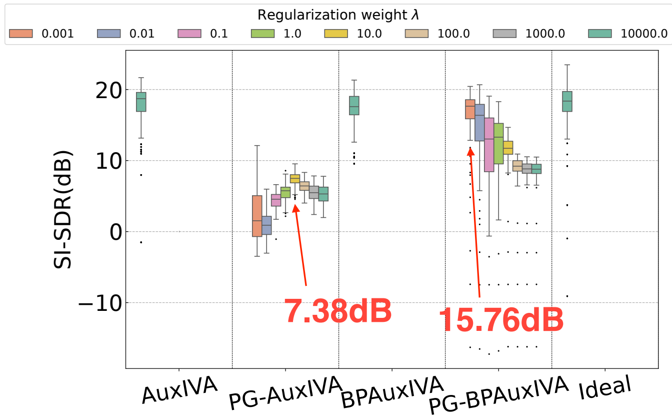
where, λ and $\boldsymbol{v}$ denote the eigenvalue and corresponding eigenvector of the matrix on the left-hand side of Eq. (29), respectively. Next, a mixing matrix was constructed by concatenating all steering vectors and normalized such that the element corresponding to the reference microphone becomes 1. The inverse of this matrix was then used as the ideal demixing matrix. For performance evaluation, we used the Scale-Invariant Signal-to-Distortion Ratio (SI-SDR) [10]. In the following, the proposed method is referred to as Prior-Guided Back-Projected AuxIVA (PG-BPAuxIVA). For comparison, PG-AuxIVA was also employed. PG-AuxIVA introduces the same prior information term as our method into the conventional cost function used for optimizing the demixing matrix before projection back. Both PG-AuxIVA and PG-BPAuxIVA utilized the same prior information, which was obtained from the ideal back-projected demixing matrix.

B. *Results and Discussions*

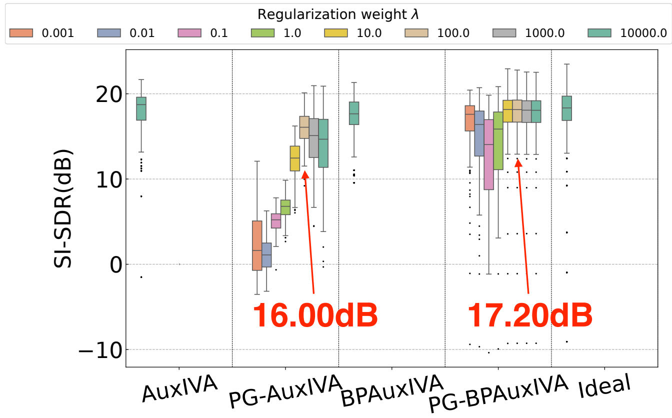
In this experiment, we compared the performance of PG-BPAuxIVA and PG-AuxIVA. To simulate a situation where the prior information is not perfectly accurate, we added noise sampled from a complex Gaussian distribution to the demixing vectors obtained from ideal steering vectors and used the resulting vectors as prior information. While this assumption may be considered near-ideal, we regard it as a valuable benchmark for assessing the potential performance in future practical applications. The Signal-to-Noise Ratio (SNR) of the vectors used as prior information was defined as the ratio of the squared L_2-norm of the clean demixing vector to that of the noise vector. We considered SNRs of 20 dB and 40 dB. In this experiment, the weight of the prior information term λ_k was set to $0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000$, which we consider a sufficiently wide range for evaluating performance. For each condition, we report the separation performance obtained with the λ_k that yielded the best performance. Note that in this experiment, λ_k was uniformly applied across all demixing vectors.

Figs. 2 and 3 show the separation performance under each SNR in two configurations depicted in Fig. 1. The red values in the figures indicate the average SI-SDR at the λ_k that yielded the best performance. A comparison of the separation performance of PG-BPAuxIVA and PG-AuxIVA at the optimal λ_k for each SNR condition showed that PG-BPAuxIVA consistently achieved higher SI-SDR in both array configurations, demonstrating its effectiveness as a method that leverages prior information of the back-projected demixing matrix. In high-SNR conditions, a larger value of λ tends to result in degraded performance, which is considered to be due to the optimization being guided too strongly by inaccurate prior information. In some cases, the best performance occurred at either the smallest or largest λ_k. When the best performance was achieved at the smallest λ_k, we observed that the performance tended to approach that of AuxIVA or BP-AuxIVA ($\lambda_k = 0$). Note that in some cases, depending on the value of λ, the performance degraded even when prior information was introduced. We consider this to be because adding the prior term changes the shape of the objective function, so that the algorithm may converge to local minima, or the optimum of the modified objective may not correspond to the point that yields the best separation performance.

To investigate the effect of prior information on convergence, we further analyzed the convergence behavior of the objective function by comparing PG-BPAuxIVA, AuxIVA, and BP-AuxIVA. In this experiment, λ_k was tested across values $0.001, 0.01, 0.1, 1$. This limited range was chosen because for larger values of λ_k, the convergence behavior showed no significant changes visually in the plotted results. For fairness, the prior information term of PG-BPAuxIVA was excluded from the objective function calculation, and the AuxIVA objective was used for all methods. As shown in Fig. 4, PG-BPAuxIVA converged faster than BP-AuxIVA. Furthermore, increasing λ_k further improved the convergence speed, highlighting the utility of the prior information. While the convergence speed varied depending on the source, PG-BPAuxIVA occasionally outperformed AuxIVA as well.



(a) SNR = 20 dB



(b) SNR = 40 dB

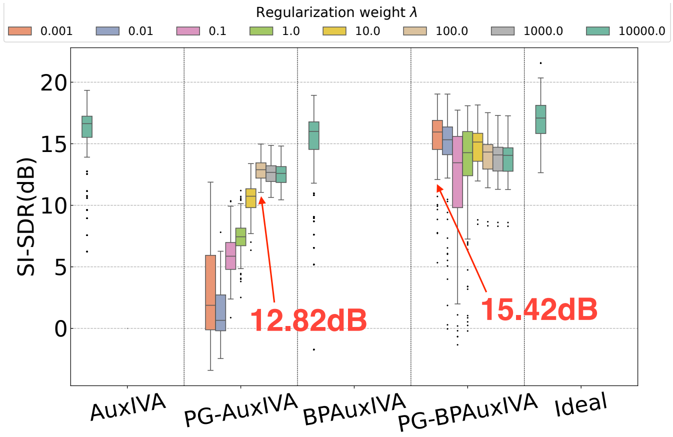
Fig. 2: Separation performances under each SNR in the centralized configuration Fig. 1a.

VI. Conclusion

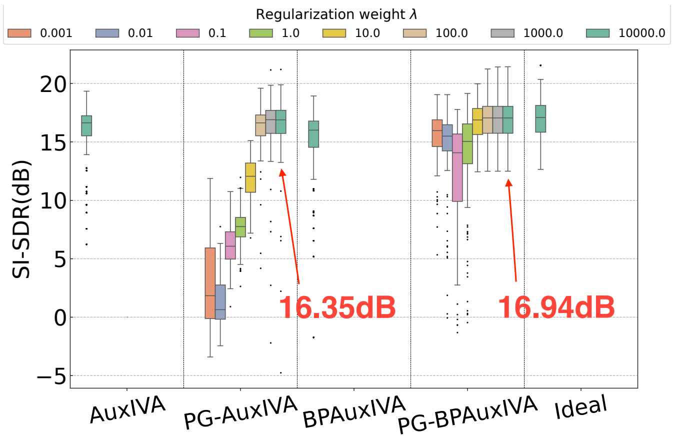
In this paper, we proposed a method that incorporates prior information of back-projected demixing vectors into the update of the back-projected matrix. Simulated experiments with conventional and other approaches that utilize prior information demonstrated that the proposed approach achieves superior separation performance. Although the experiments were conducted under the assumption that the ideal demixing vectors are partially known, future work will explore the use of more flexible prior information—such as known steering vectors—to further enhance the performance of blind source separation systems.

Acknowledgment

This work was supported by Grant-in-Aid for Scientific Research Grant Numbers 25H01150.

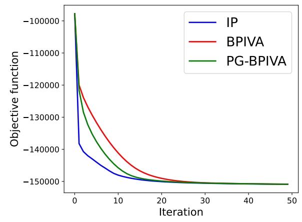


(a) SNR = 20 dB

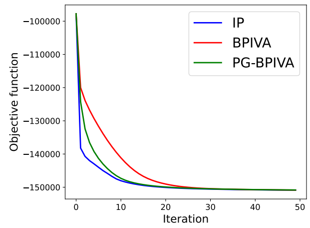


(b) SNR = 40 dB

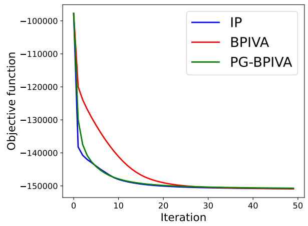
Fig. 3: Separation performances under each SNR in the centralized configuration Fig. 1b.



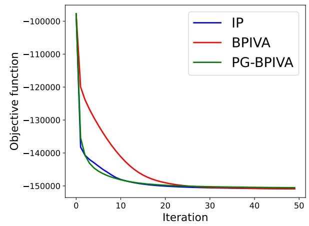
(a) $\lambda = 0.001$



(b) $\lambda = 0.01$



(c) $\lambda = 0.1$



(d) $\lambda = 1$

Fig. 4: Convergence behavior of the objective function for different prior information weights.

References

[1] A. Hiroe, "Solution of permutation problem in frequency domain ICA, using multivariate probability density functions," in *Proc. Latent Variable Analysis and Signal Separation. (LVA/ICA)*, Mar. 2006, pp. 601–608.

[2] T. Kim, H. T. Attias, S.-Y. Lee, and T.-W. Lee, "Blind source separation exploiting higher-order frequency dependencies," *IEEE Audio, Speech, Language Process.*, vol. 15, no. 1, pp. 70–79, Jan. 2006.

[3] N. Ono, "Stable and fast update rules for independent vector analysis based on auxiliary function technique," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, Oct. 2011, pp. 189–192.

[4] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *The American Statistician*, vol. 58, no. 1, pp. 30–37, 2004.

[5] 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. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 746–750.

[6] N. Murata, S. Ikeda, and A. Ziehe, "An approach to blind source separation based on temporal structure of speech signals," *Neurocomputing*, vol. 41, no. 1-4, pp. 1–24, Oct. 2001.

[7] Y. Kuriki, T. Nakashima, and N. Ono, "Direct update of back-projected demixing matrices in blind source separation," in *Proc. European Signal Processing Conference (EUSIPCO)*, 2024, pp. 922–926.

[8] T. Ono, N. Ono, and S. Sagayama, "User-guided independent vector analysis with source activity tuning," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2012, pp. 2417–2420.

[9] S. Takamichi, K. Mitsui, Y. Saito, T. Koriyama, N. Tanji, and H. Saruwatari, *JVS corpus: Free japanese multi-speaker voice corpus*, Aug. 2019. eprint: 1908.06248.

[10] J. Le Roux, S. Wisdom, H. Erdogan, and J. R. Hershey, "SDR — half-baked or well done?" In *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 626–630.

[11] S. Hirata, N. Takamune, K. Yamaoka, *et al.*, "Auxiliary-function-based steering vector estimation method for spatially regularized independent low-rank matrix analysis," in *Proc. Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, Dec. 2024.

[12] T. Nakatani, T. Yoshioka, K. Kinoshita, M. Miyoshi, and B.-H. Juang, "Blind speech dereverberation with multi-channel linear prediction based on short time fourier transform representation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2008, pp. 85–88.

[13] K. Matsumoto and K. Yatabe, "Determined blind source separation using metric projection and proximity operator of log-det function under projection-back constraint," in *Proc. 15th International Conference on Sampling Theory and Applications*, 2025.

[14] 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. Audio, Speech, Language Process.*, vol. 24, no. 9, pp. 1622–1637, Sep. 2016.